

Ecstasy's after-effects

Following the retraction of a high-profile paper, the US research agency that supports research on drug abuse needs to ensure its independence from intense political pressure to prove that recreational drugs are harmful.

It was a pretty peculiar result in the first place. Neuroscientist George Ricaurte and colleagues at the Johns Hopkins School of Medicine in Baltimore, Maryland, reported that they had injected monkeys and baboons with the equivalent of three recreational hits of ecstasy, with shocking consequences. Two of the ten animals died, and nearly all the rest had damage to neurons involved in movement and mood.

It turns out that the researchers had injected the animals with methamphetamine — commonly known as speed — by mistake (see *Nature* **425**, 109; 2003). The team deserves credit for the prompt retraction of its paper in *Science*. But there remains a bad smell that the American Association for the Advancement of Science (AAAS), which publishes *Science*, and the National Institute of Drug Abuse (NIDA), which funded the research, have done little to clear up.

The reverberations of the original publication — which was subjected to far more US media coverage than this month's retraction — are considerable. The mislabelled bottle that the authors say caused their mistake contained 10 grams of methamphetamine, only about a gram of which was used in that particular study. Ricaurte told *Nature* that at least one more study will need to be retracted. The rest of the bottle's contents were used in experiments that are as yet unpublished.

The retracted paper left the public with the impression that ecstasy is far more hazardous than it may actually turn out to be. This perception may have influenced the fate of the Reducing Americans' Vulnerability to Ecstasy Act. The act, which was appended to another bill and signed into law in April, holds club owners responsible for drug use on their premises. Critics say it is unlikely to reduce ecstasy use, but may discourage club owners from voluntary measures to protect users, such as cool-off rooms for the dangerous overheating that can occur with ecstasy, as these are tantamount to admission that drug use is going on. The legislation might have passed anyway, even

if Ricaurte's study had never been published, but the news certainly lent it urgency.

The impression that low doses of ecstasy, or MDMA, are extremely dangerous — misleadingly borne out by Ricaurte's study, but not by two decades of observing the drug being used — will hamper legitimate research to determine whether MDMA could have useful psychotherapeutic properties. Proponents claim that MDMA can be a powerful adjunct to psychotherapy for conditions such as post-traumatic stress disorder, because it promotes self-awareness. At least three studies of this are now under way in Europe. But funding for such work isn't forthcoming in the United States, and proponents say that part of the blame rests with Ricaurte's now-discredited study.

Another remarkable aspect of this episode is the public endorsement of the study, at the time of its publication, by Alan Leshner, chief executive of the AAAS and former director of NIDA. It isn't clear why an officer of the AAAS should be involved at all in publicly promoting a particular result published in its journal, least of all one whose outcome was questioned at the outset by several experts. The AAAS issued the retraction late in the afternoon on Friday 5 September, resulting in low-key media coverage, which contrasts sharply with the hype surrounding the initial paper.

Some observers have in the past questioned NIDA's ability to maintain its independence in the face of the immense pressures brought to bear by those who stand behind America's interminable 'war on drugs'. Now that Leshner is at the AAAS, he needs to safeguard its independence, rather than pander to the Bush administration's jihad against recreational drug use. It falls to the new director of NIDA, Nora Volkow, to bolster NIDA's reputation. She might start with a thorough public review of the circumstances and participants' roles in one of the more bizarre episodes in the history of drug research. ■

A necessary setback for world trade

The world's poorer countries took a stand at Cancún to defend free trade in agriculture.

World trade negotiations have significant implications for many scientists. Bio-prospectors rely on international agreements if they are to scour ecosystems for interesting compounds, for example, and antiretroviral drugs will only be made available to those in need through such agreements. The future of agricultural biotechnology also rests, more or less, in the hands of the World Trade Organization.

On the face of it, the acrimonious collapse of trade negotiations in Cancún, Mexico, on 14 September, represents a setback for these efforts. Although the Convention on Biological Diversity has come into effect, as planned, and temporary arrangements are in place for the supply of antiretrovirals and other medicines to poor countries, both activities would benefit from the successful completion of the current round of trade negotiations by its target date of January 2005.

However, the basis of the collapse leaves room for guarded optimism. The talks crashed because a new alliance of developing countries and agricultural exporters, led by India and Brazil, held their

ground. They did so even as wealthy nations taunted them by proposing that new areas, such as government contracting, be opened to free competition, while leaving unaddressed the high trade barriers that exclude farmers in poor countries from rich countries' markets.

These barriers — upheld by Japan on behalf of its rice growers, by the United States for its wealthy cotton growers, among others, and by Europe for most of its agricultural sector — remain the chief impediment to free trade between nations. Every country is entitled to protect its own trade interests, but as long as these barriers remain in place, free trade remains a mirage for most of the world's population.

At Cancún, the poor countries finally said that enough is enough. This position carries a short-term cost: poorer nations will always suffer at the hands of richer ones in bilateral trade negotiations, which is why the trade talks must ultimately succeed. But for now there should be a period of silence from Western politicians on the topic of free trade, until they have garnered the courage to address their own protectionism. ■

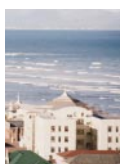




Edward Teller
Iconic nuclear-
weapons pioneer dies
aged 95
p226



On the trail
China wages war on
worms that cause
'snail fever'
p227



New wave
African institute
throws doors open to
top maths minds
p228



Splashdown
Galileo prepares to
bow out after eight
years circling Jupiter
p229

Drug companies snub antibiotics as pipeline threatens to run dry

Tom Clarke, Chicago

Major drug companies are pulling out of antibiotic development — and their timing couldn't be worse, a leading meeting on infectious disease was told this week.

Speakers at the 43rd Interscience Conference on Antimicrobial Agents and Chemotherapy in Chicago said that many firms, such as Roche and Eli Lilly, are turning away from antibiotics to concentrate on treatments for chronic illness instead.

Yet the need for new drugs has never been greater. Resistance to antibiotics is growing — 20% of infections in US hospitals involve multidrug-resistant bacteria, reports the Food and Drug Administration (FDA). Moreover, the pipeline of new antibiotics is running dry: the FDA has approved just two this year.

"There's unequivocal evidence that antimicrobial research is on a steep downward slope," said John Edwards, head of policy at the Infectious Diseases Society of America. Recent scares over diseases such as anthrax have served as a reminder that it is impossible to predict when new anti-infectious agents will be needed.

The retreat was palpable at the meeting. There were 10% fewer presentations of new



Pick 'n' mix: despite the variety of existing antibiotics, more options are needed as drug resistance mounts.

drug candidates than last year, and a slump in attendees from industry. "Attendance is down, the number of new antimicrobial agents is down," said Barbara Murray of the University of Texas Medical School in Houston, chair of the meeting's programme committee.

Big drug companies are in the financial doldrums, and antibiotics research is easy to

cut, said Steven Projan, who directs such work at Wyeth's facility in Pearl River, New York. Natural selection makes resistance inevitable, rendering any antibiotic less profitable over time. New drugs that combat resistant bacteria are often held in reserve by doctors to treat only the most stubborn infections — so they aren't big earners. And unlike drugs for chronic illnesses such as heart disease, antibiotics cure people, eliminating their customers.

The shift away from infectious diseases could have lasting effects on antibiotic research in academia, warned microbiologist Stuart Levy of Tufts University School of Medicine in Boston, Massachusetts. Jobs for microbiologists in the drug industry are becoming rarer, he said, as is industry funding for basic microbiology.

Levy suggested that antibiotics be subjected to lighter regulation and given more public-research funds, to reflect the fact that infectious diseases place a burden on society. "These are societal drugs," he argued.

But Martin Sprinklee, vice-president of product development with Bayer in West Haven, Connecticut, admitted that the industry's pleas for special treatment may receive little sympathy. "There is a perception that the industry has made unethically high returns on sales in the past," he said.

Europe offers grants to young stars

Jim Giles, London

Young researchers are being offered the chance to control their own five-year budgets of more than ≈ 1 million (US\$1 million), under a scheme unveiled this week by European research agencies.

The European Young Investigators awards are aimed at enticing the world's best young scientists to stay in Europe, or move there to work. Researchers from any country who have between two and ten years' postdoctoral experience can apply. Those confident of fulfilling the criteria — they must have the "potential to become a world-class leader" in their field — stand to receive $\approx 150,000$ –250,000 a year for five years.

The scheme is the brainchild of the European Heads of Research Councils, a body that unites the funding organizations of 15 countries. Potential applicants must first find a research institution in one of the participating countries to back their proposal. These are then vetted by the host nation's research council, before the European Science Foundation — an umbrella group of European science organizations — picks the 25 winners.

Researchers from the physical, medical, biological and social sciences are invited to apply by 15 December, and the winners will be announced next September.

► www.esf.org

Software mogul turns to mouse for genetic atlas of the brain

Rex Dalton, San Diego

Microsoft co-founder Paul Allen is contributing \$100 million to jump-start an institute in Seattle, Washington, that will seek to unravel the genetics of the mammalian brain.

The initial project at the Allen Institute for Brain Science, which was due to be announced this week, will draw up a genetic 'brain atlas' of the mouse, as a model for studying the human brain in health and disease.

"What's exciting here is the scope of this project — to have the whole genetic map done under one set of quality controls," says Larry Katz, a neuroscientist at Duke University in Durham, North Carolina. "This will be something working neuroscientists can turn to on an ongoing basis." Katz was among many leading neuroscientists who were tapped for advice during a two-year planning process initiated by Allen.

The institute will be a non-profit organization, and aims to attract more philanthropic contributions, as well as grants from federal agencies, to support its work. About 25 of the planned 75 scientific staff have already been hired.

The mouse brain atlas will map genes that are active in various parts of the brain, in a bid to gain information on the organ's development and the functions of its neurons. Institute planners say that this will help in understanding how genes influence human behaviour, as well as identifying targets for candidate drugs to treat neurological disease.

Although many neuroscientists are already studying the genetics of the central nervous system, the institute's backers say that this will be the first attempt to map the genetics of the entire mammalian brain. Up to two-thirds of the genes in both mice and humans are believed to play a crucial role in brain development and function.

Rather than pouring money into new buildings, the institute has leased premises in Seattle and will press straight ahead with its work. Allen "is interested in going forward quickly; he didn't want to wait to build a facility," explains Mark Boguski, who will direct the brain-atlas project. Boguski is a bioinformatics expert and an original member of the National Center for Biotechnology Information.

The institute's first maps are expected to be made publicly available early next year, its managers say, with the mouse brain atlas set for completion in 2006. ■

California laboratory mourns loss of H-bomb pioneer

Jonathan Knight, Livermore

The day he died, Edward Teller had been scheduled to attend a dedication ceremony at the Lawrence Livermore National Laboratory, the nuclear-weapons research complex near San Francisco that he co-founded and inspired for half a century. The ceremony was to dedicate the Edward Teller Education Center, a collaborative effort with the University of California aimed at improving science education in California schools.

The dedication of the centre to Teller, who led the team that developed the hydrogen bomb, reflects a sense at the laboratory that Teller's public image didn't always match the true character of the physicist and teacher who died on 9 September, aged 95. (A full obituary will appear in *Nature* next week.)

Teller faced ostracism from many of his peers after his infamous 1954 testimony against Robert Oppenheimer, the first director of the Los Alamos National Laboratory in New Mexico and father of the atomic bomb. Questioned about whether Oppenheimer should still be allowed access to highly classified work, Teller told a congressional committee: "I feel that I would like to see the vital interests of this country in hands which I understand better, and therefore trust more."

These fateful words helped to strip Oppenheimer of his security clearance, ruining his career — and permanently alienating Teller from other physicists. "People who were his old friends and colleagues turned their backs on him," says Ray Kidder, a retired weapons scientist who worked for Teller at Livermore. In the 1980s, Teller again clashed with other physicists, forcefully advocating the Star Wars missile-defence systems proposed by President Ronald Reagan.

But last week at Livermore, those mourning Teller were stressing other aspects of his life's work. Teaching was his passion, they say, and until shortly before his death, he spent many hours at local high schools where he hoped to inspire young minds.

Teller helped to found the Livermore laboratory to develop the hydrogen bomb — an idea that many physicists, including Oppenheimer, had rejected on technical or ethical grounds. He actually directed the lab for only two years, from 1958 to 1960 — the administrative aspects of the position didn't suit him, colleagues say. But right to the end, he helped to guide the lab by maintaining close relationships with subsequent directors and senior scientists. "You always paid attention when he had a strong idea," says Bruce Tarter, who directed the lab from 1994 to 2002.



Edward Teller at Livermore last year: his influence at the lab continued to the end.

"He had so many more interesting ideas than anyone else."

Teller's collaborative spirit, dedication to critical thinking and drive to find the best solution to any given problem left a permanent impression on generations of Livermore scientists, says John Nuckolls, a protégé of Teller's and the lab's director from 1988 to 1994. "I look at this laboratory as Teller's legacy in the twenty-first century," he says.

Colleagues recall a man who was extremely focused. "Obsession is a word that comes to mind," says Marshall Rosenbluth, a student of Teller's who went on to become a leading plasma physicist. But Rosenbluth also recalls a mentor "full of charm and charisma" who was always keen to lend support — and publication credit — to his students.

Kidder recalls travelling with the lab's fourth director, John Foster, for a meeting at Teller's home in Berkeley. "The mountain was coming to Mohammed," he says. "Teller stayed in the background, but he had enormous influence over what was going on."

Despite the effects of age — he used a wheelchair and his eyesight was fading — Teller spent two days a week at the lab right up until the end. The week before he died he was there, discussing cosmology and nuclear physics with colleagues.

And although he will no longer roam its halls, Teller will leave an indelible mark on Livermore for years to come, Nuckolls predicts. "His personality was embedded in the lab," he says. "I have a great sense of loss, but it's going to go on." ■

Additional reporting by Geoff Brumfiel in Washington.

B. MARGOT/AP

Worm genes harnessed to tackle snail fever

David Cyranoski, Shanghai

Geneticists are taking some early steps towards a fresh strategy to combat 'snail fever', a potentially fatal liver disease that affects more than a million people in China.

A team led by scientists at the Chinese National Human Genome Center in Shanghai this week published details of some 13,000 genes of the parasitic worm *Schistosoma japonicum* that causes the condition (W. Hu *et al. Nature Genet.* doi:10.1038/ng1236). Researchers now hope that an analysis of the parasite's genome will help to combat its spread.

The worm lives in freshwater snails found chiefly along the Yangtze river in central China. According to Chinese researchers, snail fever, or schistosomiasis, was fairly prevalent until the 1950s, when rudimentary public-health measures helped to curb its spread. But it has recently been making a comeback, and in July Chinese vice-premier Yi Wu ranked it next to severe acute respiratory syndrome (SARS) and AIDS as one of China's main public-health challenges.

Other species of the worm, most notably *S. mansoni*, infect humans, mainly in South America and Africa. The World Health Organization estimates that up to 200 million people in more than 70 countries are infected with *Schistosoma*.

The worm passes from its host into fresh water, from where it can penetrate human skin and infect various organs. About 40 days after entering the body, it produces eggs in the liver that can block blood flow there.



A life of slime: freshwater snails can carry a parasite that causes liver disease in people.

Symptoms include fever, stiffness of limbs and an enlarged liver and spleen. Up to 65 million people living in tropical areas of China are at risk of contracting the potentially fatal disease, public-health officials say.

Almost 50 years ago, China's government mobilized tens of thousands of peasants to take steps to reduce the number of people infected, including collecting and burying snails in soil. But recently the number of cases has increased again. The World Health Organization has cited construction of the Three Gorges Dam on the Yangtze as a possible factor in the disease's revival.

Schistosomiasis can be treated with various drugs, but these are expensive and do not prevent reinfection. Health officials are also

concerned that the worm may become resistant to drug treatment.

Ze-Guang Han, the geneticist who led the latest study, says that his team has identified several candidate genes encoding proteins that are specific to the worm and that could be used as the basis for a vaccine to immunize people against the disease. They include proteins found on the parasite's cell membrane and those involved in egg generation. "The development of a vaccine is critical for many developing countries," Han says.

According to the researchers, their work also offers some insight into the worm's unusual life cycle. Early on, for example, the male and female parasites are separate entities, but as adults the female lives tucked inside the male. "Biologically, this is a unique organism for study," says Guo-Ping Zhao, the Shanghai centre's deputy director.

The study shows that many receptor proteins on the surface of worm cells, which respond to signalling molecules from other cells, are similar to those found in the worm's various hosts. This partly explains how the worm hijacks its host's signals to spur its own proliferation and how it evades detection by the host's immune system. "We can really understand the interplay between the host and parasite," says Han.

The Shanghai centre now plans to sequence the parasite's entire genome, which is estimated to contain up to 270 million base pairs. Researchers in the United States are already leading an international effort to sequence the genome of *S. mansoni*. ■

Chemists seek image overhaul

Helen Pearson, New York

The annual meeting of the American Chemical Society (ACS), held in New York last week, saw chemists address perhaps the biggest problem facing their discipline: why young people are turning their backs on it.

The number of US bachelor's degrees awarded in chemistry fell by 8% between 1997 and 2001, society officials said. If the trend continues, "we could put ourselves out of business", warned environmental chemist Alan Elzerman of Clemson University in South Carolina.

Speakers told a seminar on 9 September that the discipline's best prospects lie in reinventing the undergraduate chemistry curriculum. The same lacklustre lectures and textbooks have been lulling students to sleep for decades, they said.

The seminar followed on from a three-day meeting in Washington DC in June, which brought together around 50 chemists

and science educators. And its conclusions echoed those of the June meeting, with participants agreeing that one reason that courses don't attract students is chemistry's generally poor public image. Chemistry departments also fail to advertise overlapping, 'sexy' disciplines such as nanotechnology or proteomics, they added.

The National Science Foundation (NSF) has ploughed \$15 million–20 million over the past ten years into initiatives aimed at revamping the chemistry undergraduate curriculum, but the effort needs even more investment, said Eli Pearce of the Polytechnic University in New York, a former ACS president.

But some scientists claim that the ACS's own membership is too old to relate to young people's doubts about chemistry. Nearly 40% of its 163,000 members are over 51 — compared with 11% aged 30 or below. Some seminars "look like something out of a



retirement community", says 31-year-old Darrell Coleman of drug firm Eli Lilly at Research Park Triangle, North Carolina.

Sylvia Ware, an ACS official working with its education committee, says that some of its members aren't facing up to the issue. "I think we underestimate the level of effort needed," she says. "There won't be major change until the grass roots say we need it." ■

African institute ready for meeting of mathematical minds

Tom Clarke, Johannesburg

More than 100 academics from around the world have volunteered to lecture at a pan-African postgraduate mathematics institute that opens its doors in Cape Town, South Africa, this week.

The African Institute for Mathematical Sciences (AIMS) will fill an important gap in Africa's education system, says Neil Turok, a cosmologist at the University of Cambridge, UK, and one of the institute's founders.

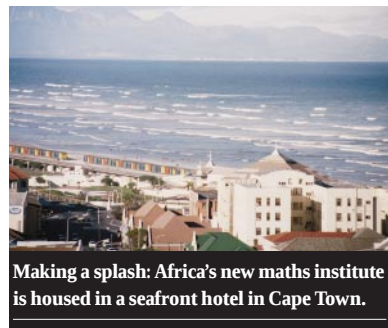
AIMS will recruit teaching expertise from around the world to provide free postgraduate diploma courses, which will include options in topics such as bioinformatics and fluid mechanics, as well as pure and applied mathematics.

"I hope to gain the knowledge to allow me to contribute to the development of my country and Africa as a whole," says Justin Bazimaziki from Rwanda, one of the first 30 students from 15 countries to win a place at the institute.

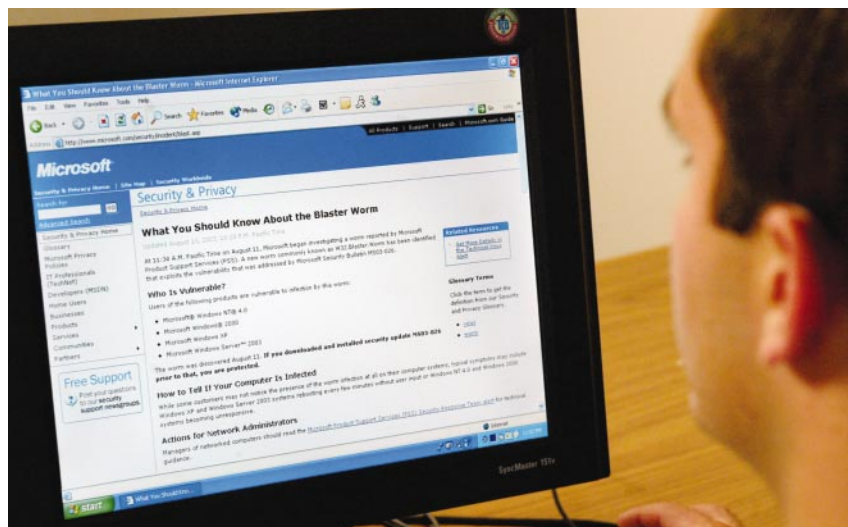
AIMS has not been expensive to set up, Turok says. It will house students and staff in a donated former hotel at the unfashionable end of Cape Town's seafront. The library is stocked with donated books, and lecturers from universities worldwide will give 4–8 weeks of their time to teach classes.

The International Council of Scientific Unions — a confederation of the world's main scientific societies — has contributed US\$150,000 to help get the project off the ground, and charities, corporations and the South African government will pay the institute's estimated \$500,000 annual running costs.

The institute's founders are aware that its graduates may choose to leave Africa once their training is complete. But Bazimaziki, who travelled for a week to Cape Town to enrol, says that he, for one, will stay. "I don't know at what level I can contribute to my country's development. But I'm going to make an effort," he says. ■



Making a splash: Africa's new maths institute is housed in a seafront hotel in Cape Town.



Cast off: academic researchers fear that they will be frozen out of a Pentagon computer worm project.

Academics sidelined in battle against computer worms

Declan Butler

The US Defense Advanced Research Projects Agency (DARPA) has decided to classify a major research programme aimed at building defences against computer worms. The move has angered scientists who argue that both universities and the military would benefit if some of the project's results were published openly.

The \$30.4-million programme is intended to counter the rising threat of worms that could hit every vulnerable machine on a network within seconds (see *Nature* 425, 3; 2003). It aims to develop techniques that will automatically identify new worms and quarantine them before they can spread.

Initial discussions on the programme included the protection of civilian computer networks, according to several participants. But when it was announced in March, the project was restricted to military networks, although elements of it were to be unclassified, enabling university researchers to participate. DARPA has now decided to classify the entire programme.

Stuart Staniford, president of Silicon Defense, a computer-security firm in Eureka, California, says he appreciates the need to classify the design details of secure systems. But he criticizes the decision to classify the whole programme, pointing out that this will exclude academics, as most of them don't have security clearance to do classified work. Those who do have clearance are not allowed to do the work on campus.

Several academics echo his complaints. "The growth of unclassified research in worm technology will be slower than it would otherwise have been," says one, who did not want

to be named. One DARPA official privately agrees that it is "a challenge" for the agency to fund academic research in classified areas, and that it tries to "identify long-term problems that can be researched in an unclassified setting by the academic community".

DARPA's director, Tony Tether, hinted at greater classification of computer-security research in testimony given on 14 May to the House Committee on Science. He said that although DARPA's research in the area had proved useful in both commercial and military systems, the focus now had to be on "specific problems the Department of Defense needs solved for network-centric warfare".

But one scientist at a leading US university, who wants to remain anonymous, says: "Once you start classifying, it shuts down that field of inquiry." The worm programme, he says, contained "tons" of basic research that could have remained unclassified.

Researchers say that fresh, unclassified awards for computer security at DARPA have almost dried up, and that other agencies which might support the work — such as the Department of Homeland Security and the National Science Foundation (NSF) — have yet to step into the void. Congress has approved a measure that would allow the NSF and the National Institute of Standards and Technology to spend \$880 million over five years on computer-security research, but this needs to be passed by appropriations committees in Congress before it becomes available.

Carl Landwehr of the NSF's computer-science directorate points out that DARPA's primary concern will always be military needs, adding that "if they find it appropriate to classify, they will". ■

Flight veteran embraces NASA's science wing

Tony Reichhardt, Washington

John Grunsfeld, NASA's new chief scientist, isn't merely passionate about space science: "I've been willing to risk my life to do science," he says. From others, that statement might sound boastful or delusional. But it is true enough in Grunsfeld's case — he has travelled on the space shuttle four times, including a 16-day research flight in 1995 and two trips to upgrade the Hubble Space Telescope.

Grunsfeld took over this month as NASA's fourth chief scientist since former administrator Dan Goldin created the position ten years ago. Since then the job has carried little political clout and no budgetary authority. But these are turbulent times at the space agency, with major decisions about its future goals looming large on the horizon. Grunsfeld — an astronomer who succeeds two biologists in the position — might just be the man of the moment, observers of the space agency say.

He brings to the job a solid research background. A 44-year-old native of Chicago, he obtained a doctorate in physics from the University of Chicago and did postdoctorate work there and at the California Institute of Technology (Caltech) in Pasadena, before beginning his astronaut training in 1992.

A specialist in γ -ray and X-ray astronomy, he has worked at the Very Large Array in Socorro, New Mexico, as well as on spaceborne instruments such as the Hubble and the Compton Gamma Ray Observatory. He has published his work in places such as the *Astrophysical Journal* — no mean feat considering the time demands of shuttle training.



Space man: John Grunsfeld's own missions have convinced him that human spaceflight is vital.

Anneila Sargent, an astronomer at Caltech, says that researchers should find Grunsfeld easy to talk to. "I see John as approachable," she says, adding that he talks scientists' language. As chief scientist, he will be responsible for advising NASA administrator Sean O'Keefe on the agency's entire research portfolio — not just astronomy but also Earth science, as well as the biological and materials-science experiments planned for the International Space Station.

Grunsfeld says that he is well aware of scientists' scepticism about the space station

(see *Nature* **418**, 263; 2002), but thinks that it is unfair to judge its usefulness while it is still under construction. He says that critics of the station may have the wrong idea about it because of their own backgrounds: laboratory scientists, he points out, are used to working in fully controlled environments. "Field scientists have a more accurate idea of how much scientific return we can get," he says, as they are used to unpredictability and to handling logistical challenges. Doing science in space, he contends, is more akin to "an athletic event".

His two successful Hubble missions are his proudest achievements as an astronaut, Grunsfeld says, adding that the telescope-servicing operations were a 'poster child' for cooperation between NASA's astronaut programme and its scientific projects. Furthering such endeavours, he says, will be "absolutely top of my list" of priorities as chief scientist. In the past year he has been involved in planning for future space telescopes that could be serviced by a combination of humans and robots — just the sort of human-spaceflight objective that NASA badly needs in the wake of the Columbia accident.

Such projects will need large infusions of money, as well as advocates like Grunsfeld, if they are to come to fruition. If NASA pursues an ambitious programme of human exploration, says Joe Alexander, director of the National Academy of Sciences' Space Studies Board, Grunsfeld's talents as an astronaut and astronomer could prove to be a useful combination. ■

Galileo gears up for swansong as crash-landing looms

John Moore

NASA's most successful planetary probe, Galileo, is due to end its mission on 21 September by crash-landing into Jupiter. There is just enough propellant on board to point its antenna towards Earth, so NASA scientists hope to retrieve more data in the craft's final days.

Galileo's lifetime has been extended three times by NASA since it arrived at Jupiter in December 1995, helping it to collect more information than any other planetary mission to date. Its observations led to the discovery that Jupiter's moon Europa probably has an ocean beneath its icy surface, and that Io harbours a giant volcano. In 1994, Galileo was close enough to the planet to see comet Shoemaker-Levy 9 break up and pummel the surface, giving scientists a better idea of what might happen in a similar collision with Earth.

If left alone, Galileo could have stayed in



Giant finale: Jupiter will shortly swallow Galileo as the prolific probe plummets to its end.

orbit around Jupiter for the next 60 years. But there was a risk that the craft would accidentally crash into and contaminate Europa, which planetary scientists think might harbour life. Intentionally crashing a spacecraft and getting data along the way isn't unprecedented — Magellan was

smashed into Venus in 1994, for example.

In Galileo's final 24 hours, it will be trained on Amalthea, one of Jupiter's inner satellites. Researchers think that it may have a comet-like structure, shedding pieces that could make up one of Jupiter's faint rings.

If Galileo is still transmitting data in its final hour, scientists hope to determine where the outer edge of Jupiter's atmosphere begins. At present, no one knows how far the planet's gases extend beyond its surface, says Galileo's project manager, Claudia Alexander.

But it is not clear whether Galileo's instruments will still be functioning by then. Researchers expect the craft to be torn apart in the atmosphere on its speedy descent at 50 kilometres per second. Details of Galileo's fate will never be known, as its final few minutes will be spent on the far side of the planet, out of sight from Earth. ■

Mixed results for conservation efforts leave species dying

Washington Two snapshots of the Earth's biodiversity released this week highlight both good and bad news for the planet — great strides have been made to protect the environment, but hundreds of threatened species have been left out in the cold.

The United Nations Environment Programme and the World Conservation Union together report that protected areas have expanded to nearly 19 million square kilometres, an area larger than China and covering 12% of the planet's land surface. This is a huge improvement on the first Parks Congress in 1962, when just 2.4 million square kilometres were protected.

But many of the sites are too small or isolated to be effective in conserving species, says a second study by Conservation International of Washington DC. It shows that more than 700 birds, mammals and amphibians on the World Conservation Union's Red List of Threatened Species remain completely unprotected. The situation may be even worse in the oceans. Both results were presented at the 5th World Parks Congress in Durban, South Africa.



The next to go? Yellow-eared parrots are precariously perched on the brink of extinction.

Prizes aplenty honour genetics and astronomy

New York Two top science prizes, behind only the Nobels in prestige, were awarded last week: the Laskers, for medical research, and the Balzans, this year devoted to infrared astronomy and evolutionary genetics.

The Lasker Award for basic research went to Robert Roeder of Rockefeller University in New York, who reproduced the transcription of DNA into RNA in a test tube and opened up the process for study. The clinical award was shared by Marc Feldmann and Ravinder Maini of Imperial College London, for their discovery of a treatment for autoimmune diseases such as rheumatoid arthritis. An award for public service went to the paralysed actor Christopher Reeve for his advocacy of medical research.

One Balzan Prize went to Reinhard Genzel, a director of the Max Planck Institute for Extraterrestrial Physics in Garching,

Oops! Sorry, we've dropped your satellite

Washington A \$239-million US weather satellite was dropped onto its side at the Lockheed Martin plant in Sunnyvale, California, last week, doing an unknown amount of damage to the craft. The NOAA-N Prime spacecraft, which was scheduled to go into polar orbit in 2008, was being shifted from a vertical to a horizontal position when it slipped from the cart and crashed to the floor. An initial assessment revealed that 24 bolts were missing from the cart — it is thought that engineers took them



out a few days earlier without properly recording their actions. Both Lockheed and a government team from NASA, the National Oceanic and Atmospheric Administration and the US Air Force have begun investigations.

Germany, for his development of telescope instrumentation and other research, some of which provided evidence for a massive black hole at the centre of our Galaxy. A second prize went to Wen-Hsiung Li of the University of Chicago, for his use of molecular clocks in studying evolution.

Bush challenged over privatization plans

San Diego The Bush administration's plan to contract out many federal jobs — including thousands of science positions — has hit a major roadblock.

On 9 September, the House of Representatives approved a bill amendment that prevents the implementation of new contracting rules that would have allowed federal jobs to be switched to private firms more easily. The legislation will now be debated in the Senate.

The amendment was championed by Chris Van Hollen (Democrat, Maryland), who represents a district that includes the National Institutes of Health (NIH). Van Hollen and others fear that contracting out scientists could disrupt research and prevent the recruitment of good scientists to agencies such as the NIH, and would save the government little or no money.

Bush administration officials say that privatization is needed to improve efficiency. President George W. Bush has threatened to veto any legislation that thwarts the move.

Merger provides stronger voice for British biologists

London Biologists in Britain are to be united under an umbrella organization launched on 15 September. The Biosciences Federation will provide its members with a single coherent voice when talking to the government and the public.

The federation is a union of two existing bodies — the Institute of Biology and the UK Life Science Committee. That gives it some 60,000 members and a broad scope, covering everything from physiology and neuroscience to microbiology and ecology. It has contributed to several reports already, including studies on the use of genetically modified crops in developing countries and the detection and dismantling of biological weapons. The organization will be similar in size and function to the US Federation of American Societies for Experimental Biology, although that is more focused on biomedical issues.

Colin Blakemore, a neuroscientist at the University of Oxford, will chair the new body initially, but plans to vacate the role when he takes over as head of Britain's Medical Research Council in October.

Heated response thaws Austria's frozen budget

Munich Scientists in Austria won a small but important victory last week, securing themselves at least some money for basic research this year.

The Austrian Science Fund (FWF), the country's only granting agency for basic research, ran into financial trouble this summer. In June, the fund received only 80% of its budget, freezing some E30 million (US\$34 million) and putting 86 previously approved projects on ice.

Austrian scientists have since been up in arms. On 11 September they presented the government with a petition signed by more than 1,500 academics, protesting against the move. Hubert Gorbach, the minister responsible for research, has now promised that half of the missing money will be provided immediately. The other half should be available by the end of the year.

Switched on to RNA

Shape-shifting RNAs that sense the environment — ‘riboswitches’ — can alter gene activity. Jonathan Knight reports on a discovery that is explaining some of the mysteries of gene regulation.

One day in 1999, molecular biologist Ron Breaker interrupted his lab’s weekly meeting to issue a challenge: to discover whether their synthetic RNA biosensors were original, or whether nature got there first.

Breaker’s team, at Yale University in New Haven, Connecticut, was crafting molecular sensors out of RNA that changes shape when it binds to a specific molecule. The team’s aim was to produce miniature devices that could, for example, monitor the biochemical secretions of living cells. These sensors were working so well that Breaker couldn’t believe that evolution had overlooked the same mechanism.

Breaker was remarkably prescient. Similar sensors, it turned out, had been staring biologists in the face for nearly 30 years. In an eye-opening series of experiments, Breaker’s team has revealed that these RNA sensors, dubbed ‘riboswitches’, represent a previously overlooked mechanism of gene regulation. “We can explain a lot of the unexplained findings that have appeared in the literature over the years,” Breaker says.

Until Breaker’s team revealed its results, biologists had assumed that the regulation of gene activity in response to environmental cues was mediated only by proteins. In the classic model of gene regulation, cells monitor their environment through a variety of specialized sensor proteins deployed either on their surfaces or internally. For example, a protein in the bacterium *Escherichia coli* detects the presence of the amino acid tryptophan, allowing *E. coli* to shut down its own tryptophan production when a ready supply of this essential nutrient exists, thus avoiding waste.

Detective work

Sensor proteins typically trigger signalling pathways, a series of protein interactions that results in a protein binding to regulatory sequences of DNA that control genes — as the tryptophan detector protein of *E. coli* does. Binding of the protein to DNA influences the activity of another group of proteins that controls the expression of genes to produce messenger RNA (mRNA). In the case of tryptophan production in *E. coli*, these mRNAs code for the enzymes that manufacture the amino acid.

Riboswitches are mRNAs that sense the environment directly, shutting themselves



Mimicking nature: Ron Breaker (left) and his team are developing synthetic RNA switches.

down in response to particular chemical cues. For example, bacterial genes for enzymes that direct the synthesis of vitamin B12 use a riboswitch¹. The mRNAs transcribed from these genes fold into a specialized shape that creates a binding pocket for coenzyme B12 — the activated form of the vitamin. When coenzyme B12 lands in this pocket, the mRNA shifts its shape in such a way as to mask a nearby sequence that would otherwise tell the cell’s protein-manufacturing factories, known as ribosomes, to ‘start reading here’. When the coenzyme B12 is abundant, these sequences are hidden, and enzymes for B12 synthesis are no longer produced.

Seven types of riboswitch have been found in bacteria so far. These include switches controlling the manufacture of the vitamins B1 and B2 (refs 2–4) and the nucleotide guanine⁵. By searching databases of nucleotide sequence, Breaker has found candidate riboswitches in plants and fungi⁶. “Higher organisms might be riddled with riboswitches,” he says.

Some of the biosynthetic pathways in which riboswitches are now implicated have

been studied intensely for nearly three decades. So why were these control elements missed? One reason is that techniques for working with RNA were not good enough until recently. But another has to do with a long-standing prejudice against mRNA: for years it was considered as nothing more than a carrier of genetic instructions.

Role reversal

In the 1980s, however, researchers learned that certain RNA molecules, dubbed ribozymes, could catalyse biochemical reactions, a job previously thought to be the exclusive province of enzymes, which are proteins^{7,8}. Later, two research groups independently found that RNA could behave like another type of protein: antibodies. The teams synthesized RNA molecules — now known as aptamers — that, like antibodies, can latch on tightly to specific target molecules^{9,10}.

Researchers then began to speculate about the existence of aptamers in living organisms. “We used to say it would be surprising if nature didn’t take advantage



Life's rich pattern: organisms as diverse as rice (left) and the bacterium *Escherichia coli* have riboswitches.

of this," says Gary Stormo, who studies RNA at Washington University in St Louis, Missouri. Several years later, this is exactly what Breaker found: a riboswitch is a specialized type of aptamer in which the act of binding to its target alters a gene's activity.

Unbelievable

Many biologists initially resisted the idea of natural aptamers, because it conflicted with traditional notions of what RNA can do. "I still give talks where scientists walk out of the room shaking their heads," says Larry Gold, who led one of the teams that first made synthetic aptamers while he was at the University of Colorado at Boulder. "The entire idea of aptamers seems completely unbelievable to them."

Nevertheless, aptamers have useful applications. Gold is now chief science officer and chairman at SomaLogic, a company in Boulder that is developing aptamer-based proteomic 'chips' to study the profile of

proteins produced in particular cells or tissues.

Other companies are exploring the use of aptamers as drugs. Breaker stumbled upon natural riboswitches while working on a biotech application for synthetic aptamers. And in May 2001, he and Andrew Ellington of the University of Texas at Austin helped to found Archemix, a company based in Cambridge, Massachusetts, that is developing ways to use riboswitches in research and medicine. One application couples an aptamer to a ribozyme so that the ribozyme is only active when the aptamer binds to its target. Ellington is also interested in using riboswitches in gene therapy, allowing patients to take pills to switch introduced genes on or off.

More fundamentally, some researchers believe that natural riboswitches represent an evolutionary throwback to a time near the dawn of life, before DNA and proteins, when RNA reigned supreme. In this world, RNA would have handled both the storage of genetic information — DNA's job in

most organisms today — and its expression, which is now the role of proteins. RNA-only genomes still exist in some viruses. And ribozymes are now known to catalyse a slew of important biochemical reactions. What's been missing from the 'RNA-world' hypothesis is a mechanism whereby organisms react in precise ways to their environment — which is where riboswitches come in.

What might riboswitches have controlled in the ancient RNA world? After all, there was no protein synthesis for them to block. Breaker suggests that riboswitches might have controlled ribozymes directly — as do the RNA sensors that he has designed. A nucleotide might have shut down the ribozymes that catalysed nucleotide synthesis, for example.

Back to the future

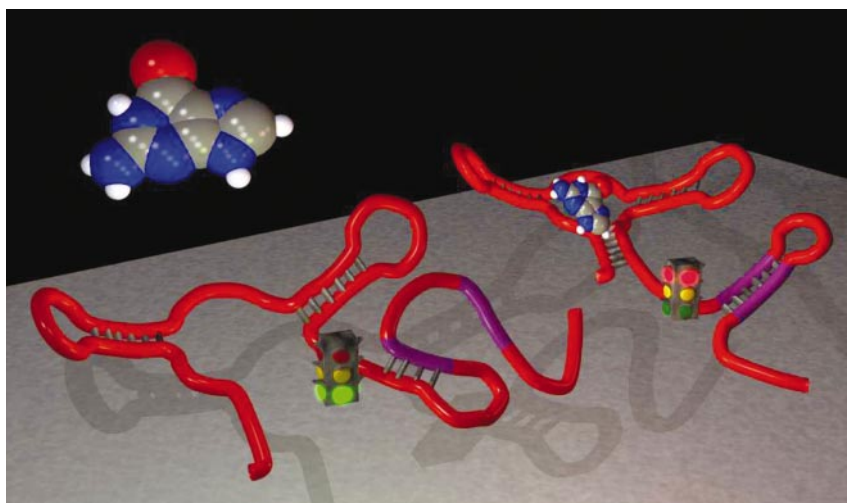
Other researchers are not convinced that Breaker's riboswitches are the direct descendants of sensors from the RNA world. Ellington agrees that riboswitches played a role in early evolution, but he suspects that since then they have emerged several times over the history of life on Earth. If modern riboswitches were direct descendants from the RNA world, he argues, they would be much more widespread.

Perhaps riboswitches are more widespread than we realize. Although sequence comparisons have so far turned up candidates in only a few species of plants and fungi⁷, the number discovered could be limited by our knowledge of what riboswitches look like. If they are very common, their sequences could be as varied as the molecules they bind to, says Evgeny Nudler, who heads a team working on riboswitches at New York University in Manhattan.

The big question now is whether riboswitches represent an occasional oddity, or a widespread and biologically important mechanism of gene control. If they start turning up in less ancient biochemical pathways throughout the plant and animal kingdoms, biologists will certainly sit up and take notice. Gerald Joyce, a ribozyme researcher at the Scripps Research Institute in La Jolla, California, agrees, "If we found them in a real blood-and-guts metabolic pathway, that would be really incredible."

Jonathan Knight writes for *Nature* from San Francisco.

1. Nahvi, A. *et al. Chem. Biol.* **9**, 1043–1049 (2002).
2. Winkler, W., Nahvi, A. & Breaker, R. R. *Nature* **419**, 952–956 (2002).
3. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. *Proc. Natl Acad. Sci USA* **99**, 15908–15913 (2002).
4. Mironov, A. S. *et al. Cell* **111**, 747–756 (2002).
5. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. *Cell* **113**, 577–586 (2003).
6. Sudarsan, N., Barrick, J. E. & Breaker, R. R. *RNA* **9**, 644–647 (2003).
7. Kruger, K. *et al. Cell* **31**, 147–157 (1982).
8. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. *Cell* **35**, 849–857 (1983).
9. Ellington, A. D. & Szostak, J. W. *Nature* **346**, 818–822 (1990).
10. Tuerk, C. & Gold, L. *Science* **249**, 505–510 (1990).



Turned off: in the absence of its target molecule (left) this riboswitch gives gene expression the green light to go ahead, but when the molecule binds, expression is halted.



Spare me the lecture

US research universities, with their enormous classes, have a poor reputation for teaching science. Experts agree that a shake-up is needed, but which strategies work best? Kendall Powell goes back to school.

Richard McCray says he feels like Oprah Winfrey, running up and down the lecture hall with a microphone, mediating student discussion. A professor at the University of Colorado at Boulder, McCray began turning his introductory astronomy course inside out three years ago.

Rather than lecturing to 200-plus students at a time, McCray divides them into 'cooperative learning teams' of about a dozen people, throws problems at them over the Internet, and then uses the lecture hall to discuss their various solutions. He did not innovate for the sake of it — he was deeply worried about the poor teaching performance of America's leading research universities. "We're losing talented people; we're driving students away," he says. "They don't like these large lecture courses."

Although educational institutions in many countries are struggling with similar issues, US research universities face a particularly tough problem in matching their specialist interests to the demands of society. A typical introductory science class consists of

a couple of hundred students, most of whom have no plans to continue with the subject, nor a great deal of prior knowledge. But in a world that increasingly depends on science and technology, it is more important than ever that these students learn the scientific basics. And that makes McCray's bleak assessment of the teaching failure of US research universities a matter of genuine social concern.

Some science professors are now trying to put new spins on their old curricula — using interactive computer technology, for instance, and employing various tactics to get students to discuss their ideas. Others, with an eye on the future, are trying to turn today's science undergraduates into tomorrow's high-school teachers (see 'Those who teach, learn', overleaf). And a handful of educational innovators are applying the scientific method to their teaching experiments, testing new methods to determine whether they bring about improvements in learning.

Transforming the traditional learning environment is tough, particularly in a system where teaching excellence is not generally

rewarded with career advancement. What's more, some education experts argue that the scientists who are now developing an interest in the subject are ignoring prior research, and are in danger of reinventing the wheel.

But everyone agrees that the standard 'lecture-then-test' format is failing — particularly where lectures are delivered to huge numbers of students at a time. Evidence of this failure is provided by assessments such as the Force Concept Inventory (FCI), a multiple-choice test designed to examine students' understanding of Newton's laws of mechanics. Developed around a decade ago by David Hestenes, a physicist turned education researcher at Arizona State University in Tempe, the FCI has changed some researchers' opinions of their teaching techniques. When he first heard about the FCI, applied physicist Eric Mazur of Harvard University in Cambridge, Massachusetts, assumed that his élite students would perform perfectly well in the traditional lecture setting. So when they received an average FCI score of 70, where 80 is considered a pass, he got "a slap in the face".

As the number of students in each class is unlikely to fall any time soon, thinking of a way to make a 200-person course seem more student-centred tops the list of desired reforms. The solution advocated by McCray, who chairs the US National Research Council's Committee on Undergraduate Science Education, is to ask small groups of students to work on problems together. In the lecture theatre, he calls on each team to read its answer aloud so that the entire class can discuss it. In addition to these *Oprah*-style debates, McCray's classes sometimes resemble the 'ask the audience' segment of television's *Who Wants to be a Millionaire*? Class members have electronic clickers so that McCray can take a problem that has stumped many teams, turn it into a multiple-choice question, and ask everyone to vote on it. This instant feedback allows him to see if a common misconception has tripped up his students — and if so, to dispel it rapidly.

Stretching exercise

Not all students enjoy the 'show' — many complain about the lack of a lecture and almost all find themselves outside their comfort zone. "I hated McCray's class when it first started," says Awwad AlWassem, now a fourth-year journalism student at Boulder. But he says that he appreciated the experience of working with classmates in a mock professional setting. He also says that he remembers more from the astronomy course than from other, lecture-based courses: "When someone tells you directions, it's not the same as if you've driven it already."

Mazur now uses similar methods to McCray's. "I did it out of total despair," he recalls. "I felt that nobody had understood what I had just taught." As an alternative to

lectures, he tried asking students to interrogate each other. Impressed with the results, he went on to base classes around such peer instruction, where students struggle with a problem, predict the answer, and then try to convince their neighbour of their argument (C. H. Crouch and E. Mazur *Am. J. Phys.* 69, 970–977; 2001). Mazur also designed ConceptTests, sets of qualitative exam questions that rely on understanding a concept rather than simply using physical formulae. His methods have been adopted by physics teachers around the United States and have also been adapted for chemistry, astronomy, geology and mathematics courses.

Another innovation involves the use of information technology, which frees up more time for interactive discussions. Many instructors now combine texts and lecture notes into a single online hypertext document, often with animated illustrations and links out for more information. Once this basic instruction has been moved online, class time can be devoted to questions and discussion — in many cases using voting by electronic clickers, as favoured by McCray.

Online teaching tools can also go beyond lecture notes. When Carl Wieman, a physicist at the University of Colorado at Boulder, shared the 2001 Nobel Prize in Physics for his work on ultracold atoms, he decided to channel some of the prize money into his Physics Education Technology project. Wieman devised simple lab exercises, which programmers turned into software designed to help students visualize what would otherwise be abstract concepts. One of the programs shows the build-up of static electricity when students rub the foot of dancer 'John Travoltage' across a carpet. Another gets

students to build a circuit using a battery, switch, lightbulb and resistors. When the switch is thrown, the program illustrates the flow of electrons. "These allow students to visualize concepts in the same way that a trained physicist does," Wieman says.

Lecture lag

Although many physics instructors were initially sceptical of the new techniques, there is evidence that they work. Wieman says that test scores have gone up by an average of two grades since he introduced his computer aids. And a survey that tested more than 6,000 students with the FCI at several higher-education institutions showed that straight lectures — whether boring or entertaining — proved significantly less effective than more interactive courses (R. R. Hake *Am. J. Phys.* 66, 64–74; 1998). Other factors, such as class size and student preparedness, had little or no influence. Such results have spread like wildfire among physics departments concerned with improving teaching, says Mazur.

Next, say experts, studies need to be done to compare different types of interactive teaching. Daniel MacIsaac, a science education researcher at Buffalo State College in upstate New York, has developed a tool that may help with this. The Reformed Teaching Observation Protocol is a 25-question survey that is filled in by teachers observing science or maths classes. The resulting score is a rank, from 0 to 100, of the basic effectiveness of the innovations in promoting learning. Tests such as the FCI can then be used to see whether specific concepts are getting through to the students.

Tools similar to the FCI are also proving the worth of alternative teaching techniques in other areas of science. Mike Zeilik of the University of New Mexico in Albuquerque has developed the Astronomy Diagnostic Test (ADT). He found that women score more poorly on the ADT than men when taught in traditional lectures. But Zeilik says that when he replaced lectures with an interactive course in which students constructed 'concept maps' — diagrams linking key ideas into their proper relationships — the gender gap was closed, and both men and women did better overall. In unpublished work, MacIsaac has also shown that he can close gender and minority gaps on the FCI by using 'whiteboarding', in which students sketch out problems and present them to the class on shared whiteboards.

Another new test — the Biology Concept Inventories (BCI) — could also shake up biology professors in the same way that the FCI and ADT prodded physicists and astronomers from their complacency. "You can't fool yourself that because you lectured, it was learned," says Michael Klymkowsky, a developmental biologist at the University of Colorado at Boulder. To create the BCI, Klymkowsky and his Colorado colleagues survey professors in a given discipline to find out which key ideas should be covered. The next step is to interview



Well rounded: Daniel MacIsaac (centre) encourages students to form interactive discussion groups.

D. MACISAAC/B. AHERN/AZTEC 2002

students to find out what misconceptions they have about these topics. These are then used to form alternatives to the correct answer on the final, multiple-choice test. Ultimately, the researchers hope to create a tool that covers the whole of biology — however daunting that may sound. “I’ve bitten off 400 times more than I can chew,” jokes Klymkowsky.

For those who have been studying science education for many years, the influx of high-profile scientists is good news. “Welcome aboard! I want them all,” enthuses Diane Ebert-May, a plant ecologist at Michigan State University in East Lansing who has been working on biology education since 1987. But there is also concern that newcomers are ignoring existing knowledge and expertise or, worse still, being unscientific. “You have to think hard about your question and research design,” says Ebert-May. An education-research question poses greater challenges than an investigator’s own disciplinary research, she says, because it uses human subjects and few controls, and deals with the slippery data of learning.

Prior knowledge

“If I were going into condensed-matter physics, I would certainly feel obligated to find out what people in that field already know, read the literature, and see what experiments have already been done,” says Lillian McDermott, director of the Physics Education Group at the University of Washington in Seattle, one of the oldest science-education research programmes in the United States. McDermott and others say they encourage anyone interested in improving science education, but demand the same rigour in testing new methods as in any other discipline.

McCray, for instance, has drawn fire from education researchers for not using the ADT or another established tool to measure the

When someone gives you directions, it's not the same as if you've driven it already.

Awwad AlWassem

success of his astronomy course. Now that the course has developed over many years, McCray says he will this year begin to test student learning, using the Student Assessment of Learning Gains tool, an online student survey to measure how well an instructor’s teaching methods have helped students learn.

Zeilik says that many instructors are not willing to use the ADT and other proven assessment tools because they fear that they will not pass inspection, or that testing will reveal gender or other bias in their courses. And Ebert-May warns that some newcomers to the field are in danger of wasting their enthusiasm on experimental teaching projects that largely repeat what has gone before. “Does someone need to test peer instruction again? No, we know it works and now we’ve moved on to more sophisticated things,” she says.

But scientists, including McCray and Wieman, complain that some current assessment tools do not properly test whether students grasp concepts in the same way that an expert would. What’s more, they say, science-education research must strive not only for learning gains, but also for methods that are practical, given the resources of research faculties. “This is a problem with much education research,” says Wieman. “They produce models for teaching as if time and cost are no object.”

This culture clash between scientists and

science-education researchers may settle into a peaceful collaboration as more scientists join the movement. Ebert-May agrees that assessments need to move beyond simple multiple-choice tests in order to be able to measure critical scientific skills. She and others say the most important goal is to raise awareness among researchers that interactive, enquiry-based instruction works better than lecturing. With enough experts giving their advice on scientific content, assessment tools will improve. Conversely, by drawing on research that shows which classroom techniques make the grade, learning will hopefully improve, too.

McCray says the two groups have begun to build bridges, and he is confident that the new teaching methods will catch on, partly because it is becoming more difficult for research universities to cling to the traditional introductory courses designed to filter out the scientists from the English majors. “We can’t stay in this state where it doesn’t matter if you learn anything,” says Klymkowsky. Instead, says McCray, teachers should start reaching the majority, not just the minority of scientifically inclined students who will succeed no matter how the class is taught. “We need to reach that middle group of students who can become scientifically literate.” ■

Kendall Powell is a science writer in Broomfield, Colorado.

Richard McCray’s Astronomy 1020 course

♦ cosmos.colorado.edu/astronomy

Physics Education Technology project

♦ www.colorado.edu/physics/phet

Project Galileo

♦ galileo.harvard.edu

Biology Concept Inventories

♦ bioliteracy.net

Astronomy Diagnostic Test & Force Concept Inventory

♦ www.flaguide.org

Physics Education Group

♦ www.phys.washington.edu/groups/peg

Those who teach, learn

For many people, science literacy begins and ends in childhood — which is why some researchers who are exploring ways to improve undergraduate science education are also trying to turn their students into tomorrow’s school teachers.

Mike Marder, for instance, co-directs the UTeach programme, a collaboration between the Colleges of Natural Sciences and Education at the University of Texas at Austin. The four-year programme enrolls roughly 180 students each year, who finish with both a science degree and a teaching certificate.

“These are typical students,” says Marder. In the beginning, he says, they have a fairly shaky grasp of the subject matter. “That’s just not

acceptable for a teacher.” So Marder and his colleagues have students design their own experiments, give presentations, prepare lessons and eventually go to local schools. They also meet with ‘master teachers’ — former high-school science teachers who help them perfect their classroom technique.

Marder says the programme’s success is not just in the number of science teachers it has produced, but also in the number and type of students it draws to science more generally. The percentage of UTeach students who stay in their science major subject is double that of the university’s general science majors. The programme also attracts and retains more women

and minorities, both of which are typically underrepresented in the sciences. UTeach has 64% women and 26% Hispanics, African-Americans and Native Americans, whereas the College of Natural Sciences has just over 50% women and a minority population of 17.5%.

At the University of Washington in Seattle, students who wish to earn a certificate to teach high-school physics must take Physics by Inquiry, a course that exposes them to the same pitfalls that their future students are likely to meet. Even those who majored in physics must take the course, says Lillian McDermott, director of the university’s Physics Education Group. “They go through that relearning process to get them

closer to the way they should be teaching,” she says.

For students who are unsure if teaching is in their future, Richard McCray’s learning assistantships at the University of Colorado give top candidates a taste. The programme is one of 18 funded by a Science, Technology, Engineering, and Mathematics Teacher Preparation grant from the National Science Foundation to encourage bright science scholars to consider a teaching career.

K.P.

UTeach

♦ www.uteach.utexas.edu

NSF Science, Technology, Engineering and Mathematics Teacher Preparation

♦ www.ehr.nsf.gov/duel/programs/stemtp

It's time for the US and Muslims to work together

Each side has problems to solve and much to gain from contact with the wider world.

Sir — After 11 September 2001, the United States adopted precautionary measures to protect its people and institutions. Many institutions have slammed their doors to Muslim scientists. As a scientist and writer interested in science and society, I am concerned that the current political and emotional upheavals will severely handicap the growth of science in the Islamic world.

Discriminatory US policies discourage US institutions from recruiting Muslim scientists. These policies include the US Department of Agriculture's ban on scientists from several countries including traditional Islamic areas; extensive background checks on Muslims; and the recent law requiring men from Islamic countries to report to immigration offices.

These exclusion policies cover a broad range of science, from agriculture to IT, and will severely inhibit the growth of science in the Islamic world. To avert catastrophe, all groups — Muslims,

Americans and Europeans — must act.

Muslims must recognize the need to spend petrochemical wealth in establishing centres of scientific excellence and creating an environment conducive to scientific growth in all Islamic countries (not just Arab ones). Muslim nations must formulate relevant R&D programmes, restructure institutions to end corruption and dynastic practices, and collaborate with both Western and newly industrialized states to nurture a strong science base. Finally, Muslims must demand more from their leaders, so that they can become equal to other societies.

US organizations and individual scientists must lobby hard to change post-11 September policies that are detrimental to scientific endeavour. US scientists could collaborate, mentor and otherwise assist Muslim scientists. American Muslim scientists must take the lead in Islamic science, as they live in a progressive country.

Europeans can take an active role in

bridging the rift between the US and Muslim worlds, filling the vacuum created by US precautionary policies. Europeans are already taking small steps to collaborate with the Arab countries (*Nature* **423**, 906; 2003). This process should be encouraged.

There is an urgent need to reinvigorate the rusted scientific sector of the Muslim world beyond rhetoric, marathon conferences and establishment of skeleton institutions. It requires the concerted effort of politicians, scientists, philanthropists, entrepreneurs and ordinary citizens. There is definitely room for reform and business opportunities in the Islamic world.

The United States was brutally hurt by a few ruthless terrorists, and faces further threats. However, blanket repressive policies in response will cause more harm than good to the world as a whole.

Ahmed M. Dirie

California, USA

adirie2001@yahoo.com

Bomb dosimetry unlikely to change risk estimates

Sir — In his News and Views article "A dose of the bomb" (*Nature* **424**, 495–496; 2003) about the impact of the new A-bomb dosimetry on radiation-risk estimates, Mark Little argues that the risks of γ -ray exposure should become more reliable. Two other factors suggest that the new dosimetry is likely to have little or no impact on risk estimates for doses that are occupationally or environmentally relevant.

The US National Council on Radiation Protection and Measurements (NCRP Report No. 126, NCRP, Bethesda, 1997) has determined the dose-rate effectiveness factor to be the main source of uncertainty in risk estimates, accounting for about 40% of total uncertainty. Overall, dosimetry uncertainties account for less than 10% of the total.

Dose-rate effectiveness factors must be applied to A-bomb-derived risk estimates to account for differences in temporal dose delivery. The bomb exposed people to very high dose rates (essentially instantaneous exposure), whereas risk estimates are usually applied to occupational or environmental situations where the dose is delivered at a significantly lower rate and risks are assumed to be lower because of biological repair mechanisms.

The second source of uncertainty is dose extrapolation. A-bomb risks are based on excess cancer deaths in people exposed to doses greater than 200 mSv (D. A. Pierce

et al. Radiat. Res. **146**, 1–27; 1996). Below this dose the total excess number of cancer deaths is too small to be used reliably in risk estimation. Accordingly, risks calculated from high-dose data are extrapolated using the linear non-threshold theory to predict risks at small doses. Assuming that occupational and environmental doses are around 2 mSv per year, dose extrapolation is significant and the theoretically derived risks are highly uncertain.

Kenneth L. Mossman

School of Life Sciences, University Office of Radiation Safety, Arizona State University, Tempe, Arizona 85287-3501, USA

Mark Little replies — I am grateful for Dr Mossman's comments on my article. As I stated in the penultimate paragraph, sources of uncertainty other than the dosimetry have to be taken into account in deriving risk estimates from the A-bomb data. Constraints of space did not allow me to go into these in detail. I would concur with Dr Mossman that extrapolation of risks to low doses and low dose rates is one of the more substantial of these, others being the extrapolation of risks to the end of life and across populations, for example from a Japanese to a UK population (NCRP Report No. 126, NCRP, Bethesda, 1997).

The various sorts of bias (selection, ascertainment and so on) to which epidemiological studies are prone should also be considered. As discussed in my article, selection bias in A-bomb survivors may be significant, and as with the problems

of extrapolation of dose and dose rate, may largely invalidate the A-bomb risk estimates. However, as noted previously, cancer risks derived from A-bomb survivor data are statistically consistent with those observed in groups exposed at moderate-to-low doses and dose rates (M. P. Little *et al. Radiat. Res.* **151**, 218–224; 1999; R. Wakeford *et al. Int. J. Radiat. Biol.* **79**, 293–309; 2003).

Nothing new under the Sun?

Sir — J. M. Gordon and colleagues, in their Brief Communication "Surgery by sunlight on living animals" (*Nature* **424**, 510; 2003), describe the use of a solar photocoagulator to necrose a rat liver lesion in what they believe to be "the first time that intense incoherent light has been applied successfully in an interstitial medical procedure". They may be surprised to learn that such a technique played a significant role in the development of a modern clinical ophthalmic practice: retinal photocoagulation. The German ophthalmologist Gerhard Meyer-Schwickerath (G. Meyer-Schwickerath *Ber. Dtsch. Ophthalmol. Ges.* **55**, 256–259; 1949) not only undertook the repair of retinal breaks by focused sunlight, but also designed the optical instrumentation that bears his name: the Meyer-Schwickerath coagulator.

Merrick Moseley

Department of Ophthalmology, Imperial College, St Dunstan's Road, London W6 8RP, UK

Cosmological questions

Can 'big science' shed light on dark matter and the nature of the cosmos?

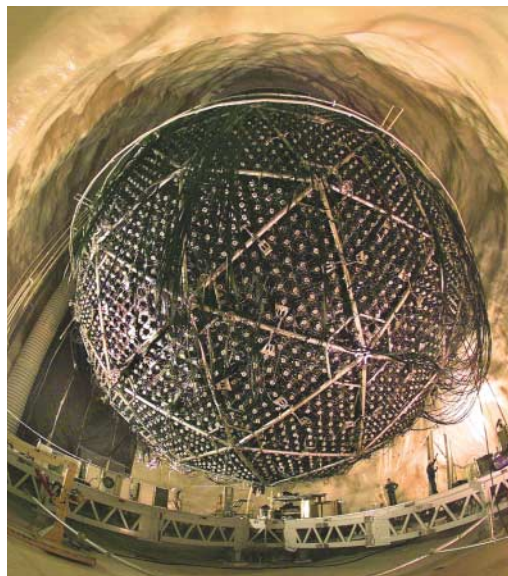
Connecting Quarks With the Cosmos: Eleven Science Questions for the New Century
by the Committee on the Physics of the Universe, National Research Council of the National Academies
National Academies Press: 2003. 222 pp.
\$39, £27.95
Martin Rees

Astronomers have traditionally been 'users' of physics, but now they are returning the compliment. They can discover new physics by probing much more extreme conditions than could ever be produced in Earth-bound laboratories: neutron stars and supernovae reach exceedingly high densities and temperatures, and cosmic rays have millions of times more energy than can be achieved in the biggest terrestrial accelerator. There are growing links between the sciences of the very large and the very small — between cosmos and microworld. Exciting progress has already been made: recent advances in cosmology will provide an epochal chapter when the history of science in this decade is written.

Connecting Quarks With the Cosmos was written by a panel of experts, chaired by Michael Turner of Chicago University, who were charged with assessing the key challenges and opportunities at these scientific interfaces. They winnowed the list down to 11 key questions, and offered some recommendations about how to address them.

Heading the list of questions is: "What is dark matter?" Galaxies and clusters of galaxies would fly apart were they not held together by the gravitational pull of dark matter, which contributes five times more mass than the ordinary atoms that make up all the stars, planets and gas. Dark matter consists of electrically neutral particles that have survived from the hot early Universe. Identifying these particles would not only clarify what the Universe consists of, but also reveal a new type of particle. The importance of searching for such particles provides an impetus for the further development of the underground laboratories. Underground measurements of neutrinos from the Sun have already led to one of the most important discoveries in particle physics of the past 20 years — that neutrinos have a mass greater than zero.

The favouritism for matter over antimatter in the Universe could only have come about if baryon number — the number of quarks minus the number of antiquarks — were not absolutely constant. The downside of this non-conservation of baryon number is that protons can decay. (The instability



Looking for clues: the Sudbury Neutrino Observatory in Canada provides insight into the nature of the Universe.

of protons is another of the key questions considered.) This incredibly slow process — maybe one decay per year in a thousand-ton tank — is another candidate for an experiment in an underground laboratory.

In the past few years an even more baffling mystery has emerged: 70% of the cosmic mass-energy seems to be neither in atoms nor in dark matter, but latent as 'dark energy' in empty space. Associated with this energy is a negative pressure; according to Einstein, this then exerts an 'antigravity', causing cosmic acceleration. The nature of the dark energy — another of the questions tackled in this book — can be pinned down by more careful measurements of the relationship between redshift and distance, a task that requires telescopes that can detect large numbers of distant supernovae.

Another question is: "Did Einstein have the last word on gravity?" Einstein has been impressively vindicated by high-precision tests within the Solar System, by gravitational lensing and by pulsar studies. These tests, however, pertain to the 'weak field' limit, where Einstein's theory only slightly modifies Newton's. Einstein predicted that black holes should be standardized objects, characterized by just two numbers — mass and spin — and described precisely by equations first described by Roy Kerr, but we have no direct evidence that this is so. To test strong gravity, we need to detect (and model) gravitational waves from black holes crashing together, and study the radiation from gas swirling in the strong-field region, or energy extracted from the spinning hole itself.

We know that Einstein's theory is transcended deep inside black holes, and also at the beginning of the Universe, where conditions are far more extreme than any extrapolation of physics that we can test.

Are there extra spatial dimensions beyond the three that we are familiar with? According to string theory, ten dimensions may have played a role in the ultra-early Universe — but if so, why did only three expand? Cosmological observations, particularly of the microwave background radiation, may offer clues. Perhaps new physics is needed not only where space-time curvature is large, but when particles move very fast. The best probes for such an effect would be the most energetic cosmic rays, whose speed differs from that of light by just 1 part in 10^{22} . The Auger array in Argentina, now under construction, should be able to collect enough

data to settle some issues that have been tantalizingly uncertain for several years.

The focus of this book is 'big science' — spacecraft, accelerators and telescopes. But it is gratifying that some fundamental questions can still be tackled with tabletop experiments. For instance, ingeniously designed torsion balances can test the gravitational inverse square law on scales of less than a millimetre, constraining the scale on which extra spatial dimensions could be wrapped up.

The book reflects an admirable feature of the organization of big science in the United States. Long-term priorities are set by panels whose members are selected for scientific distinction and balance, and who engage in wide and open consultation. Such consultation has value in itself, broadening the perspective of individual researchers, and forging a firm consensus. Set up by the National Academies, the panels operate at arm's length from funding agencies, such as the National Science Foundation, NASA and the Department of Energy. Even so, their reports are taken very seriously by the funding agencies and by the congressional committees that sanction individual projects.

Given its collective authorship, one would expect this book to be authoritative, and it is, but it is also attractively readable. That such an excellent book can come from a committee testifies to the intellectual thoroughness with which US priorities in big science are optimized. ■

Martin Rees is at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

From Manhattan to Los Alamos

The National Labs: Science in an American System 1947–1974

by Peter J. Westwick

Harvard University Press: 2003. 384 pp.

\$51.50, £33.50

Herbert York

The National Laboratories are a group of ten research and development institutions, sponsored and largely funded by the US Department of Energy and managed in a style known as GOCO: government-owned, contractor-operated. The individual labs employ up to 10,000 people, including scientists and engineers from a wide range of disciplines, and have annual operating costs ranging from a few hundred million dollars to more than \$1 billion. People who are able — and rash enough — to make such judgments believe that two them, Los Alamos and Lawrence Livermore, are, in whatever order, perhaps the two best laboratories in the world.

The laboratories have a common root in the Manhattan Project in the Second World War. The project was set up to exploit the newly discovered process of nuclear fission with a view to developing an atomic bomb. The immediate purpose of this, in turn, was to bring about a quick and decisive end to the deadly series of wars that had plagued the world since 1914. The project was directed and funded by the US Army Engineer Corps and implemented by several separate contractors, including pre-existing university laboratories at Berkeley, Chicago, Columbia and Ames; newly created labs at Los Alamos and Clinton; and totally new production facilities of unprecedented size, built and operated by large private companies at Oak Ridge and Hanford.

After a general demobilization and a brief hesitation, growing concerns about the Cold War led to a decision to continue with the development of nuclear weapons under the management and direction of a new civilian-led government agency, the Atomic Energy Commission. The programme was quickly expanded to include the development and operation of giant particle accelerators, commercial power reactors, and research in biology and medicine. Most of it was done in facilities adapted or built for the wartime project, but some new laboratories, notably Brookhaven and Argonne, were established in the late 1940s.

The intensifying Cold War led to new thinking about what the United States really needed, and the labs were quick to recognize and exploit new opportunities. These factors stimulated them to evolve and expand, something they have done almost continu-



Brave new world: work on the Manhattan Project to make the atomic bomb led to the first National Laboratories.

ously ever since. Some programmes have been discarded along the way (nuclear propulsion for aircraft, ramjets and rockets among them), but many more new ones have been instituted. These include the construction and operation of new accelerators and 'advanced' light sources; work on nuclear fusion and other potential sources of energy; energy conservation; environmental studies; huge lasers for fusion and isotope separation; unraveling of the human genome; the development of new methods for gathering and analysing intelligence on foreign weapons programmes; and the creation of a variety of means for coping with weapons of mass destruction in the context of terrorism.

The laboratories have also been major drivers in the development of modern computers by being extremely demanding customers with lots of money and smart people. In effect, they subsidized the exuberant expansion of the US computer industry and the resulting enormous computational capability that the world now enjoys.

In addition to their obvious contributions to national security and economic development, the labs played other less well known but nonetheless important roles in the international arena. International cooperation in basic science comes as no surprise, but there has even been close cooperation in the area of weapons development and related technologies. The working relationship with British weapon designers that existed during the war years was revived, with some limitations, in the 1950s. Parallel but more guarded cooperation with the French dates from the 1960s. And when the Cold War ended, US and Russian weapons laboratories began a major two-way programme of exchanges and visits. A similar, although smaller and more tightly circumscribed, relationship also

exists with Chinese weaponeers.

Westwick describes these events clearly, accurately and (mostly) comprehensively. There are other books on individual labs, but this is the only one that pulls them all together. Historians of science and of the Cold War, as well as all students of bureaucracy, will find this book useful, and staff and alumni of the system — by now a very large audience — will find it an enjoyable read.

One of Westwick's main points is that the interactions between the labs and with the government produced a system that is much more than the sum of its component parts. The labs simultaneously compete and cooperate with each other. The government micromanages the labs, while allowing them remarkable freedom and flexi-

bility in some areas. The core programme of each laboratory is highly specialized but is surrounded by many other broadly based activities. Westwick coins the term 'systemicity' to describe these processes and their results.

As well as the direct impact of their programmes, the labs have also strongly influenced the course of national and world events in ways not covered in the book. These include lab alumni who served as high-level officials. Two have been cabinet members — Harold Brown was defence secretary, and Glenn Seaborg was chairman of the Atomic Energy Commission — and several dozen more have filled influential and authoritative positions in national government departments, or served as US representatives to international organizations such as NATO and the United Nations. Many others — employees as well as alumni — have participated in high-level advisory committees dealing with national security, including arms control and intelligence.

Finally, many of the best-known alumni have been key participants in public-interest groups, including those engaged in peace activism. The international Pugwash movement was founded largely by veterans of the Manhattan Project. At the national level, two peace and disarmament advocacy groups (the Federation of American Scientists and the Council for a Livable World) were founded by alumni, and a constant stream of lab personnel and alumni have testified before the Congress, both in support of and in opposition to government policies and decisions. ■

Herbert York is director emeritus of the Institute on Global Conflict and Cooperation, University of California, San Diego, La Jolla, California 92093, USA.

Life, the Universe and everything

Magic Universe: The Oxford Guide to Modern Science

by Nigel Calder

Oxford University Press: 2003. 720 pp.

£25, \$40

Nicholas Wade

Nigel Calder is a journalist of great skill and experience, and his craft is evident in this collection of essays about a broad range of scientific topics. The essays, 120 in all, are deft vignettes that lay out the principal scientific issues in a field, portray the leading players, and sketch the recent development of the subject. Each essay is cross-referenced to others, loosely interweaving the main themes of the book: physics, biology and cosmology. It is this interconnectedness, the author explains, that gives the Universe the magic referred to in the book's title.

The entries in this guide to modern science are arranged alphabetically, essentially randomizing the subject matter. Turning to the entry on the human genome, I was impressed at the author's ability to compress many different kinds of information into a readable and insightful narrative. Much British coverage of the race to sequence the genome has tended to portray Craig Venter as an American profiteer, slapping patents on the genome against the principled opposition of Britain's John Sulston, who insisted that everything be in the public domain.

Calder has picked up on the fact that the ideology concealed a more personal issue. Venter, an academic at heart, was forced to seek private funding after being denied adequate support by the US National Institutes of Health. It is an irony, too deep to explore here, that the Sanger Centre near Cambridge, UK, where Sulston was based, was able to do its pioneering work on the nematode and human genomes only because its patron, the Wellcome Trust, reaped a windfall, under the wicked patent system, from the profits accrued by Burroughs Wellcome on the sale of drugs such as AZT.

Calder also notes that it was only after being denied support from the National Institutes of Health that Kari Stefansson set up deCODE Genetics in Iceland to search for haplotypes associated with common diseases. That two scientists of such abundant talent and drive as Venter and Stefansson should be obliged to work outside the official funding system could be a matter of serious discomfort to those persuaded of the general merit of allocating funds by peer review.

But the book's format of an A-to-Z guide proves rather restrictive. In attempting to cover both the genome-sequencing efforts and deCODE Genetics' search for haplotypes



Exhibition

Building on nature

Organic architecture was espoused by Frank Lloyd Wright as long ago as 1908, and there were stray examples of natural influences on buildings before that. Joseph Paxton's design for the Crystal Palace, for example, which was built in London in 1851, was inspired by the ribs of the giant *Victoria regia* lily, and the splayed base of the Eiffel Tower in Paris copies the buttress principle of large tree trunks. But the great flowering of organically inspired architecture is happening now.

Hugh Aldersey-Williams has brought together models, sketches and photographs of buildings — displayed alongside stuffed examples of the animals that inspired them, borrowed from the neighbouring Natural History Museum — in a new exhibition at London's Victoria and Albert Museum.

Sometimes the inspiration is iconic. The image used in publicity for the exhibition is

Santiago Calatrava's Milwaukee Art Museum (shown above), which mimics a giant bird (or, to some eyes, the tail of a humpback whale). Sometimes the inspiration is structural. David Marks and Julia Barfield's competition-winning design for a cafe in the redeveloped Bullring shopping centre in Birmingham, UK, is based on a nautilus shell and the Fibonacci series, the mathematical justification for the ancient golden section, or ideal sensual proportion.

Architectural firms whose work is included in the exhibition include Foster and Partners, Gehry Partners, Nicholas Grimshaw and Partners, Renzo Piano Building Workshop, Ushida Findlay and Wilkinson Eyre. Some of the entrancing hybrid structures showcased in the exhibition are as yet unbuilt — the exhibition should encourage this situation to change.

Peter Forbes

The Zoomorphic exhibition runs at the Victoria and Albert Museum, London, until 4 January 2004.

in this single entry, Calder ends up giving no detailed picture of either.

An entry on 'prehistoric genes' turns out to be an account of the dispersal of the human population from its ancestral African homeland. I felt that this essay did not do complete justice to its material and seemed less up-to-date than the genome entry.

The essay on 'life's origin' discusses comets and the RNA world, but not the theory recently proposed by Günter Wächtershäuser. He believes that the beginning of life should be sought not in the cell or in nucleic acids, but in a natural metabolic cycle that perpetuated itself and then somehow managed to escape from its surface catalyst. In the past few years, Wächtershäuser has demonstrated

that several necessary elements of such a cycle are possible. The piece on 'nuclear weapons' shows Calder at his best: a lot of history, physics and sensible political judgement is packed unobtrusively into a breezy account of the bomb-makers and their foibles.

The book has no diagrams, which is perhaps a pity, because a great deal of scientific information can be conveyed in a graphic.

Magic Universe is a collection of generally enticing appetizers. It may be most useful for general readers, for whom it surveys many of the more interesting fields in contemporary science, leaving the reader to decide which to explore elsewhere in further depth. ■

Nicholas Wade is at The New York Times, 229 West 43rd Street, New York, New York 10036, USA.

Possible or probable?

Myles R. Allen

As a record-breaking hot summer in Europe draws to a close, the issue of global warming is once more in the news. Away from these headlines, those with the job of keeping our doorsteps dry over the coming decades, and our water, food and energy supplies secure, are now actively planning for climate change. This transition, from politics to practicalities, presents the climate-research community with new challenges. Today's coastal and water-supply engineers do not need old-style 'projections' of how the climate might respond to rising levels of the greenhouse gases, no matter how detailed. Projections of what might happen in the future are fine for lurid headlines, but practical planning needs exactly the opposite kind of information. The challenge of probabilistic — or risk-based — climate forecasting is to start saying what changes can be ruled out as unlikely, rather than simply ruled in as possible.

The Intergovernmental Panel on Climate Change (IPCC) recognized the need for a probabilistic dimension to climate forecasting in their 2001 Assessment Report, although they refused to assign an explicit probability to their 1.4–5.8K range for projected warming over the twenty-first century — despite intense pressure to do so. One of the main reasons for the IPCC's reluctance to interpret this range as a formal uncertainty estimate was because it depends on the spread of results from a small number of climate models, included primarily because they happened to be available at the time. Models taken 'off the shelf' tend to be too similar to each other to be considered the kind of random, mutually independent sample beloved of statisticians. Checking one model's results against another's, when both have been developed with reference to the same observations, does not quite deserve the philosopher Ludwig Wittgenstein's jibe about buying several copies of the morning newspaper to assure yourself that what it says is true — it's more like buying two newspapers that rely (to an unknown degree) on the same wire service.

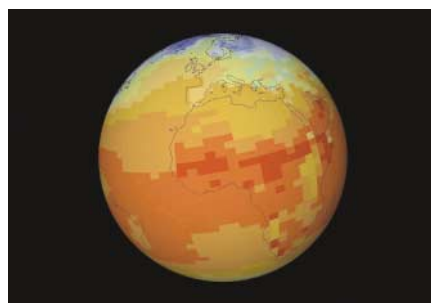
So what can be done? At first glance, generating a probabilistic climate forecast seems straightforward. Any forecast beyond the next few years must allow for uncertainty in how our models represent the climate system. This uncertainty is found in the values of crucial parameters that are not well constrained by observations — such as the reflectivity of clouds — and more generally in how processes such as cloud formation should be modelled. So we begin with

a representative sample of possible models, weight them by some measure of their similarity to the real world, and use this weighted 'ensemble' of forecasts to infer future risks.

The problem is obtaining that representative sample of possible models defined, crucially, without reference to the observations used to weight the ensemble. If the same observations are used to select models initially as are subsequently used to weight them, then we 'double-count' and inevitably underestimate uncertainties (back to the newspapers analogy). But observations are used throughout the process of climate-model development in such an ad hoc way that it is impossible to disentangle the influence of any particular data set. Leaving aside the double-counting problem, how can we be sure that our forecasts depend primarily on observations (which, although uncertain, tend to be revised and updated relatively slowly) and not on the earlier choice of models or perturbations (which are subject to the whims of expert opinion)?

To be reliable, probabilistic climate forecasts must begin with a perturbation analysis of one or more climate models to identify consistent relationships between observable quantities and forecast variables of interest. We then weight the individual perturbed model-versions to ensure that the entire ensemble accurately represents both current knowledge and uncertainty in these observable quantities. We can then infer future probabilities from the weighted forecasts, comfortable in the knowledge that — provided these relationships are consistent across physically reasonable models of varying structure and resolution — our results should depend on observations, and not a dubious earlier choice of models. It is a subtle but profound change in our attitude to climate models. Rather than providing surrogates for reality, they become tools for teasing out useful relationships between things we can observe and things we want to forecast.

This is straightforward enough when the relationships in question can be represented



World view: desktop computers are uniting to reduce uncertainty in climate simulations.

Climate forecasting

Climate modellers need to start saying what changes can be ruled out as unlikely, rather than simply ruled in as possible.

by simple, low-order functions, but most forecast quantities of interest will be related to several independent observable quantities. With realistic climate models, we not only require many simulations with different starting conditions for each model-version, but we also cannot 'aim' simulations at poorly sampled regions on the manifold. We just have to keep perturbing our model(s) until we fill out this multi-dimensional space of responses, potentially requiring hundreds of thousands of century-timescale simulations of a full-scale climate model.

Mapping the response manifold of a full-scale, non-linear climate model is a truly formidable challenge, well beyond the capabilities of conventional supercomputing resources. The only way to access sufficient resources is to use idle processing capacity on home and desktop personal computers. This is the climateprediction.net approach, proposed on these pages almost four years ago, following the successful launch of the SETI@home project (<http://setiathome.ssl.berkeley.edu>), which is now by far the largest single computation ever performed. Thanks to the support of the UK Research Councils for coupled modelling and the help of the Meteorological Office, the ingenuity of a dedicated group of scientists and the enthusiasm of a small army of beta-testers, we have configured one of the world's best climate models to run on almost any up-to-date Windows PC. If you own a PC and would like to take up the challenge of probabilistic climate forecasting, please join us at <http://www.climateprediction.net>. ■

Myles R. Allen is in the Department of Physics, University of Oxford, Oxford OX11 0QXJ, UK. This was written on behalf of the climateprediction.net project team, with assistance from D. A. Stainforth and D. J. Frame (University of Oxford), A. Kettleborough (Rutherford Appleton Laboratory) and M. Collins (UK Meteorological Office).

FURTHER READING

Houghton, J. T. et al. (eds) *Climate Change 2001: The IPCC Scientific Assessment* (Cambridge Univ. Press, Cambridge, 2001).
Schneider, S. H. *Climatic Change* **52**, 441–451 (2002)
Allen, M. R. et al. in *Proceedings of the 2002 ECMWF Predictability Seminar* 275–295 (ECMWF, Reading; online at www.climateprediction.net/science/pubs).

Wires on water

Peidong Yang

A centuries-old technique for transporting timber is the inspiration for a new method of assembling nanowires into large-scale, ordered patterns that could form the basis of a new generation of electronic devices.

In the nineteenth century, river logging became one of the principal lumbering operations in the area around the Great Lakes of North America. Logs were piled up through the winter. When the spring floods came, the logs were rolled down into the snowmelt-fed rivers, whose rapid flows carried them downstream to the sawmills. This log drive became a familiar, but still spectacular, sight, with aligned timbers covering many miles of river surface (Fig. 1). Reporting in *Nano Letters*, Whang *et al.*¹ and Tao *et al.*² have adopted a similar 'logs-on-a-river' approach on a nanoscopic level, to align semiconductor and metal nanowires on a water surface.

What do logs and nanowires have in common? They are both essentially one-dimensional objects, their diameters insignificant compared with their length. Over the past few years, one-dimensional nanostructures (whether you call them nanowires, nanorods or nanowhiskers) have attracted much attention, and considerable progress has already been made in their synthesis and their application in devices³. A formidable challenge still to be faced, however, is the hierarchical organization of these nanoscale building-blocks into functional assemblies and, ultimately, useful systems. If nanowires could be aligned and arranged into patterns, the impact would be tremendous in many areas, from nanoscale electronics and optoelectronics to molecular sensing.

Earlier attempts at nanowire alignment used microfluidic^{4,5} and electrical⁶ methods; more recently, a nano-imprinting technique using semiconductor superlattices as templates showed an impressive capability for aligning nanowire arrays of very high density⁷. But, despite some degree of success in nanowire alignment and patterning, all of these techniques fell short of achieving large-scale assembly — a milestone that has

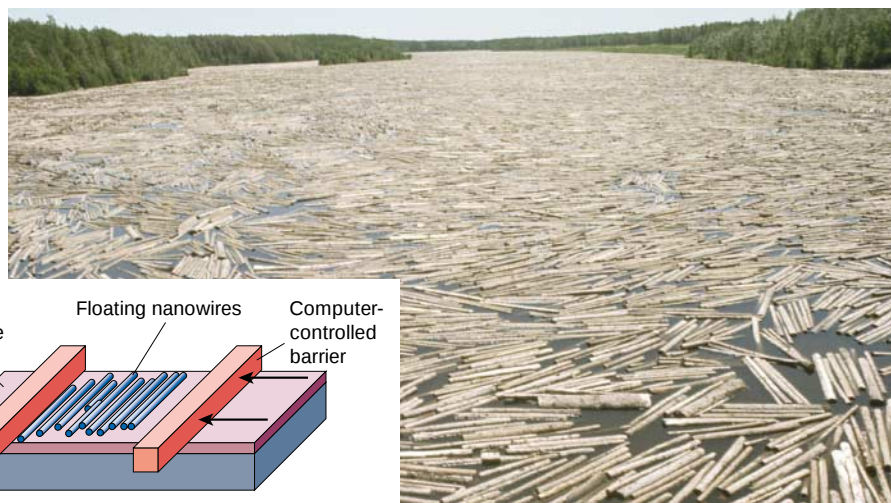


Figure 1 Logging and 'nano-logging'. Timber harvested from the forests of North America is floated downriver to sawmills. Whang *et al.*¹ and Tao *et al.*² have used a similar approach (inset) to guide the assembly of nanowires — floating the nanowires on a water surface and guiding their progress with computer-controlled barriers.

now been reached using the 'logs-on-a-river' approach^{1,2}.

Scientifically, the process is known as the Langmuir–Blodgett technique⁸. Its history can be traced back to experiments carried out by Lord Rayleigh, who found that a film of oil on water would form a layer just one molecule thick. Subsequently, Langmuir showed that monolayers of fatty acids could be compressed and made into a solid-like ordered state on the surface of water. Langmuir and Blodgett further realized that such monolayers could be transferred from the water surface onto a solid substrate by slowly passing an appropriately treated substrate through the interface between water and film.

The Langmuir–Blodgett technique has proved hugely successful in preparing monolayers of fatty acids and many other amphiphilic molecules that can be floated on the surface of water. It has been used extensively in the preparation of monolayers

for molecular electronics, and more recently to create nanocrystal monolayers with tunable properties⁹. Three years ago, my colleagues and I were the first to apply this technique to

one-dimensional nanostructures¹⁰; we succeeded in assembling short-aspect-ratio nanorods on a water surface to create textures that resembled liquid crystals (Fig. 2).

From nanocrystals to nanorods, now the Langmuir–Blodgett technique has proved equally powerful for the assembly of nanowires with much larger aspect ratios: Whang *et al.*¹ have created a system of silicon nanowires, Tao *et al.*² of silver nanowires. In both cases, suspensions of nanowires (capped or mixed with surfactants) are dispersed on the water surface of a Langmuir–Blodgett trough, the interaction between the surfactants and the nanowires causing the nanowires to float on the water surface. Then the floating nanowires are compressed to higher density on the surface, with computer-controlled trough barriers mimicking the banks of the logging river (Fig. 1). Many nanowires reorient themselves and align parallel to the trough barrier, finally forming a closely packed monolayer. This monolayer can then be transferred onto any substrate, such as silicon wafers or other plastic substances.

There are several important features of such assemblies. First, the pitch (or repeat distance) of the nanowire pattern can be controlled at nanometre or micrometre scales through the compression process. This is critical if the nanowires are to be the building-blocks of ultra-high-density microelectronics. In fact, as a first step, Whang *et al.*¹

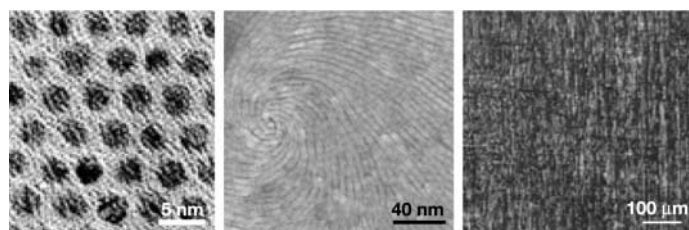


Figure 2 Langmuir–Blodgett assembly: (left to right) dots⁹, rods¹⁰ and wires¹.

have already demonstrated that these dense nanowire arrays can be interconnected with reliable metal contacts. Second, it is possible to transfer monolayers, layer by layer, to form parallel and crossed-nanowire structures that could serve as optoelectronic components.

Just as the log drive can extend over many miles of river, the Langmuir–Blodgett technique can be used to assemble large areas of nanowire monolayers on a water surface — up to 20 cm² is easily achieved. The monolayer area is limited only by the number of the nanowires dispersed on the trough surface. This type of large-scale nanowire assembly is unprecedented, and could be applied to many other one-dimensional nanostructures, including carbon nanotubes, for example. The feasibility of transferring multiple layers of metal or semiconductor nanowires onto flexible substrates also points to new directions for flexible electronics and optoelectronics.

Transforming spaghetti-like, tangled nanowires into an ordered, large-area array by the Langmuir–Blodgett technique is a remarkable feat. As Whang *et al.*¹ point out, this process offers a flexible pathway for the step-by-step assembly of virtually any nanowire material into the highly integrated and hierarchically organized nanodevices that are needed for a broad range of functional nanosystems. But this is not the end of

the story. For nanostructured technology to be competitive, being able to create high-density arrays is not enough: how to address individual elements in a high-density array and how to achieve precise layer-to-layer registration for vertical integration are just two of the many challenges still ahead. ■

Peidong Yang is in the Department of Chemistry, University of California, Berkeley, California 94720, USA.

e-mail: p_yang@uclink.berkeley.edu

1. Whang, D., Jin, S., Wu, Y. & Lieber, C. M. *Nano Lett.* **3**, 1255 (2003).
2. Tao, A. *et al.* *Nano Lett.* **3**, 1229 (2003).

3. Xia, Y. & Yang, P. *Adv. Mater.* **15**, 353–389 (2003).
4. Messer, B., Song, J. H. & Yang, P. *J. Am. Chem. Soc.* **122**, 10232–10233 (2000).
5. Huang, Y., Duan, X., Wei, Q. & Lieber, C. M. *Science* **291**, 630–633 (2001).
6. Duan, X., Huang, Y., Cui, Y., Wang, J. & Lieber, C. M. *Nature* **409**, 66–69 (2001).
7. Melosh, N. A. *et al.* *Science* **300**, 112–115 (2003).
8. Ulman, A. *An Introduction to Ultrathin Organic Films: From Langmuir–Blodgett to Self-Assembly* (Academic, San Diego, 1991).
9. Collier, C. P., Saykally, R. J., Shiang, J. J., Henrichs, S. E. & Heath, J. R. *Science* **277**, 1978–1981 (1997).
10. Kim, F., Kwan, S., Akana, J. & Yang, P. *J. Am. Chem. Soc.* **123**, 4360–4361 (2001).

Molecular biology

MicroRNA is here to stay

Philip N. Benfey

A form of gene regulation that uses small RNA molecules to bind to longer RNAs was first described over a decade ago, but was thought to be of little significance in controlling cellular processes. No longer.

The first glimpse of the wave was more than a decade ago, when a strange form of gene regulation was described that involved the binding of one RNA molecule to another¹. Then last year, with reports that there are hundreds of small RNAs in the genome^{2–5}, it came into view. There was the possibility that a whole

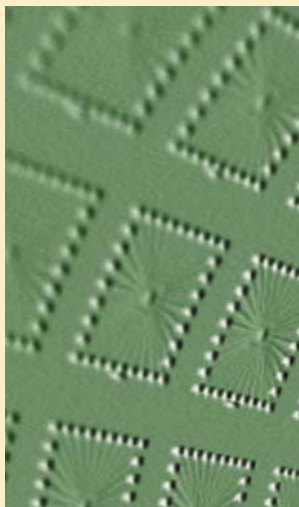
layer of gene regulation involving very small RNA molecules called microRNAs had been overlooked for 40 years. But still there were doubts as to the importance of microRNAs — the wave might yet turn out to be scarcely a ripple, let alone a tsunami. With the description by Palatnik *et al.* of the role of microRNAs in controlling plant

Electronics

Nanotechnology goes large

Discovering new ways of making and manipulating materials at nanometre scales should help to maintain the computer industry's relentless drive towards ever greater miniaturization and performance. But in this issue, Xiangfeng Duan and colleagues show that, in addition to allowing the development of high-performance nanoelectronics, these techniques may also be useful for making flexible electronics over large areas and at low cost (*Nature* **425**, 274–278; 2003).

At present, making microelectronics involves a lot of waste. More than 95% of the bulk of the precious silicon wafers from which most microchips are made serves no other purpose than as a mechanical support for the circuitry patterned into its surface. For laptop computers, digital cameras, portable music players and other high-value gadgets, the cost associated with such waste is easily absorbed into



the price. But for products such as 'smart clothing' or electronic paper — which involve higher volumes, large areas and more modest price tags — this cost becomes prohibitive. Moreover, the high temperatures required to grow crystalline silicon (in excess of

1,400 °C) make many such products difficult to produce at any price. But by growing only as much semiconductor material as is needed for electronic circuitry on the surface of an inexpensive substrate material such as glass or plastic (see picture), significant reductions in cost can be achieved.

Commercially, large-area electronic devices are based on either amorphous silicon (used in most LCD displays) or, more recently, organic semiconductors (as in the display on James Bond's electric shaver). The performance of these materials, however, is poor compared to conventional crystalline semiconductors, and is always likely to be so. But by using newly developed techniques for growing crystalline semiconductors in the form of tiny nanometre-diameter wires and ribbons — techniques that are currently being pursued for making nanoscale devices (see "Wires on water" by Peidong Yang,

above) — Duan *et al.* show that high-performance, low-cost macroelectronics could be just around the corner.

By aligning silicon nanowires or cadmium-sulphide nanoribbons between metal electrodes, the authors can create field-effect transistors — the fundamental building-blocks of modern electronic circuitry — with characteristics better than those of similar amorphous silicon or organic semiconductor devices, and approaching those of polycrystalline silicon devices. By increasing the density of wires and ribbons between the metal electrodes, Duan *et al.* expect soon to be able to improve this performance even further. And with the recent advent of nanowires made from high-mobility materials such as indium phosphide and indium arsenide, such devices could in future exceed the performance of crystalline silicon devices. **Ed Gerstner**

X. DUAN, NANOSYS INC.

development (page 257 of this issue⁶), and other new publications reporting important functions in animals^{7–9}, the wave looks very real — and big.

In the beginning there was the view that DNA makes RNA makes protein, and life was simple and good. Other functions for RNA were discovered. But they were mostly structural, such as being a part of the ribosome, the cellular site of protein manufacture. Then in 1993, while studying mutations that changed the timing of developmental events in the worm *Caenorhabditis elegans*, Victor Ambros made a startling discovery¹. A mutation that caused an increase in the translation of RNA to protein was found to be in a second, very small RNA molecule. The small RNA was shown to bind through base pairing to one end of an RNA that controlled the worm's ability to develop properly. The binding of the small RNA resulted in a block to translation of the messenger RNA. Seven years later, a second case of a small RNA used to regulate the translation of a messenger RNA was reported¹⁰. However, the small RNA was again found in worms and was controlling a similar developmental process. It was beginning to look as if this was just another baroque facet of evolution — a form of regulation that was highly specialized for one organism and one function.

What rescued microRNAs from neglect was genome sequencing. Several groups started to look for sequences within the genome that were transcribed into RNA but the RNAs were not used to make proteins. They combined bioinformatics searches with sophisticated procedures that allowed them to isolate only small RNAs. What they found was that in organisms ranging from plants to man there were hundreds of small RNAs (18–25 nucleotides) that were actively transcribed and highly conserved between related species^{2,4,5}.

The next task was to determine what these microRNAs were doing. Last year, two groups published the dramatic findings that many of the apparent targets of plant microRNAs are transcription factors implicated in the control of developmental processes^{2,4}. Transcription factors are proteins that control gene activity. So the findings led to speculation that a primary role for microRNAs in plants is to regulate gene expression after a cell division event that leads to the formation of two different cell types. But to identify putative targets through bioinformatics is one thing. To prove that microRNAs are really regulating a developmental process is quite another.

Palatnik *et al.*⁶ have done just that, although it was not the original intent of their project. They were screening a collection of mutants of the plant *Arabidopsis*, made by randomly inserting a piece of DNA into the genome that increases transcription of neighbouring genes. They found several

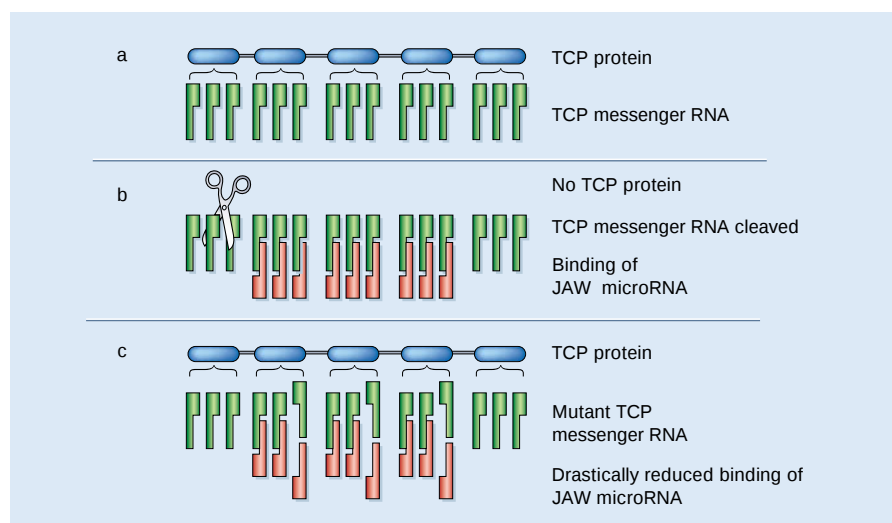


Figure 1 MicroRNAs in action. In the early days of molecular biology, RNA was known only as a way of transferring information from DNA to protein. But microRNAs do not make protein and they prevent protein from being made when they bind to messenger RNAs. **a**, When unbound, messenger RNA of the TCP gene-transcription factors makes protein. Each set of three nucleotides (green) is a codon for a specific amino acid (blue). **b**, The JAW microRNA (red) binds by complementary base pairing to the TCP messenger RNA and leads to cleavage of the messenger RNA, thus preventing protein from being made. **c**, Palatnik *et al.*⁶ changed the third nucleotide in the codons of the TCP messenger RNA. This drastically reduces the ability of the JAW microRNA to bind to its target, but does not change the protein that is made. Introducing the gene for this modified TCP messenger RNA into wild-type plants resulted in defective leaf development, indicating that microRNAs are controlling the process.

mutations that caused leaves to curl rather than lie flat. When they sequenced the region of the genome near the inserted DNA they discovered what appeared to be a gene encoding a microRNA named JAW. To identify possible targets of the microRNA, they used microarrays to compare the global expression profile of the *jaw* mutants with that of wild-type plants. Among the RNAs with the greatest differences were four members of the TCP class of transcription factors, which had been shown to control leaf curvature in snapdragon. Alignment of the JAW microRNA with the TCP RNAs showed near perfect complementarity (Fig. 1).

A series of elegant experiments showed that TCP gene function was indeed controlled by the JAW microRNA. The TCP sequence was modified so that the JAW microRNA could no longer bind while the protein made from the TCP RNA was not affected (Fig. 1). Introduction of this mutant form of the gene into wild-type plants resulted in severe developmental defects. That is, the microRNA could not carry out its job, which was evidently to control the availability of TCP protein.

Further evidence that the critical control point was exercised by microRNAs came from overexpressing a normal TCP RNA. For many transcription factors, overexpression of their RNA is like forcing too much electricity through a circuit. But overexpression of the TCP RNA caused no obvious defects. The same construct introduced into the *jaw* mutant was able to partially rescue the leaf-

curling defect. The likely explanation is that in the *jaw* mutant, the inserted DNA causes too much JAW microRNA to be produced and it is no longer restricted to certain tissues. Making more of the target RNA sops up the excess JAW microRNA.

The paper by Palatnik *et al.*⁶ provides one of the most compelling cases that the newly discovered microRNAs have an important role in controlling development. Other recently published results highlight the involvement of microRNAs in regulating processes ranging from cell proliferation and programmed cell death in flies⁷ to neuronal differentiation in humans⁸. Among the many unanswered questions that surround microRNA function a central one is, what controls microRNA expression? Is there still another level of regulation that has been overlooked — another wave just beyond the horizon? ■

Philip N. Benfey is in the Department of Biology, Box 91000, Duke University, Durham, North Carolina 27708, USA.
e-mail: philip.benfey@duke.edu

1. Lee, R. C., Feinbaum, R. L. & Ambros, V. *Cell* **75**, 843–854 (1993).
2. Llave, C., Kasschau, K. D., Rector, M. A. & Carrington, J. C. *Plant Cell* **14**, 1605–1619 (2002).
3. Lim, L. P., Glasner, M. E., Yekta, S., Burge, C. B. & Bartel, D. P. *Science* **299**, 1540 (2003).
4. Rhoades, M. W. *et al.* *Cell* **110**, 513–520 (2002).
5. Lee, R. C. & Ambros, V. *Science* **294**, 862–864 (2001).
6. Palatnik, J. F. *et al.* *Nature* **425**, 257–263 (2003).
7. Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. *Cell* **113**, 25–36 (2003).
8. Kawasaki, H. & Taira, K. *Nature* **423**, 838–842 (2003).
9. Xu, P., Vernooij, S. Y., Guo, M. & Hay, B. A. *Curr. Biol.* **13**, 790–795 (2003).
10. Reinhardt, B. J. *et al.* *Nature* **403**, 901–906 (2000).

Quantum optics

Single atom lases orderly light

Howard Carmichael and Luis A. Orozco

A laser that operates through repeated emission from a single atom is very different from the lasers we know. The beam of light produced has a more orderly photon stream than even the quietest laser.

For many physicists, the fascination of quantum systems comes from their intrinsic fluctuations — those fluctuations always present in any system as a consequence of Heisenberg's uncertainty principle. In quantum systems of practical use (such as lasers and transistors), the intrinsic fluctuations are small enough not to interfere with the function of the device. All of this changes, however, in a miniaturized version of a quantum system, when its size is shrunk down to the ultimate limit. Quantum fluctuations then dominate, and familiar devices such as lasers behave very differently.

The miniaturization of a quantum device requires better control of the elements that form it. Technical fluctuations must be suppressed if behaviour at the quantum level is to be seen. Surprisingly, the emergence of the quantum fluctuations can bring with it a paradox: as the intrinsic fluctuations become dominant, some device characteristics may fluctuate less. As a consequence — and as McKeever *et al.*¹ show on page 268 of this issue — a laser built from a single atom may emit a more orderly photon stream than a conventional many-atom system.

The basic elements of a laser are an amplifying medium (free atoms or a solid-state material) that preserves the phase of an incoming wave, and an optical cavity,

formed by two opposing partially transmitting mirrors. The mirrors force the wave to pass many times through the amplifying medium, increasing its amplification. The medium is pumped in some way (by another laser or an electric current) into an excited state from which its stimulated emission of photons into the laser mode is more favourable than spontaneous emission (which would be a net loss). The interplay creates a threshold for laser action, which begins when the gain, provided by the pumping of the medium, overwhelms the loss from spontaneous emission and mirror transmission.

Single-atom lasers call for a particularly strong coupling between the atom and the photon, as one atom must do the work normally shared by many others. Creating sufficiently strong coupling is not an easy task. For visible photons in particular, a unique set of challenges must be met, arising from the small wavelength of visible light: the atom cannot move more than a fraction of the wavelength, because any such motion introduces noise. But McKeever *et al.*¹ have overcome the technical difficulties. Although one-atom lasers (and masers, the microwave version of lasers) have been demonstrated before^{2–4}, there were always several atoms involved in creating a steady state for laser operation in these systems; the new device

is the first to operate using, as McKeever *et al.* put it, “one-and-the-same atom”. It is a model example of a system in which the quantum fluctuations have been brought to dominance.

The interaction of an electromagnetic field with a single atom has been the subject of long and fruitful theoretical studies, beginning with work in the 1960s by Jaynes and Cummings⁵, and by Paul⁶. Thirty years later, Mu and Savage⁷ explored what might happen were a laser built on this elementary interaction; they questioned popular notions, including the idea that laser light is as ‘quiet’ (free from noise) as any source of light could be. The laser (or maser) principle was first applied on the scale of single atoms in a series of experiments in which a beam of Rydberg atoms — highly excited atoms that emit and absorb microwaves — traversed a superconducting microwave cavity. The so-called micro-maser² and a similar device based on two-photon emission³ were developed. These devices achieve maser action with an extremely dilute atomic beam, ideally with no more than one atom at a time in the cavity, although many atoms must traverse the cavity to establish and sustain the maser action. A similar system operating with visible light has also been devised⁴.

McKeever *et al.*¹ have removed the final source of extraneous fluctuations by eliminating the beam of atoms. With a laser trap, they hold a cold caesium atom between the mirrors of a small optical cavity — a cavity so small that the presence of just one photon is enough to enable stimulated emission to overwhelm the spontaneous emission loss. Two lasers pump the atom, which, working alone, amplifies the light through a series of repeated stimulated emissions. Although it emits a directed beam of light, this laser is very different from the lasers we know. The threshold is absent. More fascinating, though, are the fluctuations. The light is quieter than ordinary laser light in two distinct, though interrelated, ways (Fig. 1). First, after a photon is emitted there is a characteristic delay before a second photon is likely to be emitted, with the result that the photon stream is ‘anti-bunched’. Second, the number of photons emitted in a fixed time is more predictable than for an ordinary laser (the distribution is sub-Poissonian). Overall the photon stream is more orderly than that from a many-atom laser.

Orderly photons do come at a price — a rather small flux of less than 100,000 photons per second, and continuous operation for less than one second. But new directions are opened up. In the quantum future, single-atom lasers might form circuit components for manipulating quantum information.

Howard Carmichael is in the Department of Physics, University of Auckland, Auckland, New Zealand.

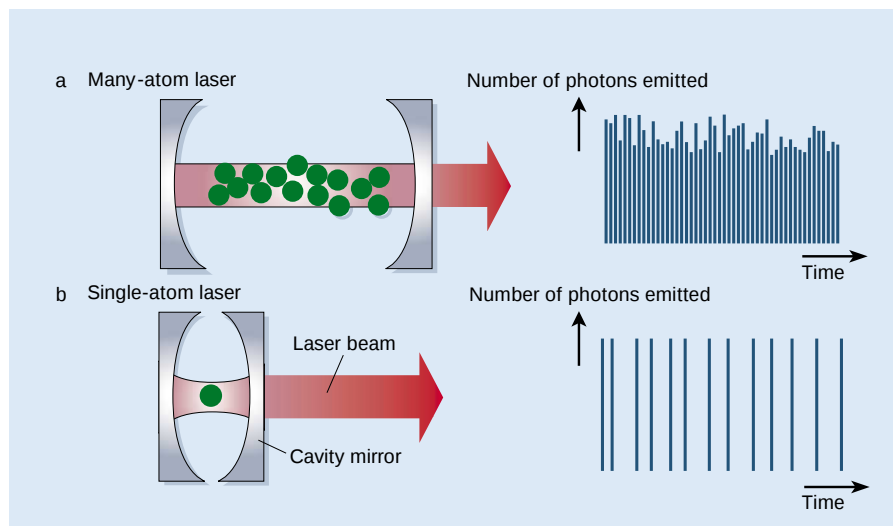


Figure 1 Photons to order. a, In a conventional, many-atom laser, stimulated emission of photons from atoms inside a mirrored cavity generates an intense beam of radiation. If the make-up of the beam is analysed over time, the flux of photons is seen to fluctuate considerably. b, In the single-atom laser demonstrated by McKeever *et al.*¹, the photons are ‘anti-bunched’, spread out over time. But the number of photons emitted over a given time interval is more predictable than for the many-atom laser.

e-mail: h.carmichael@auckland.ac.nz

Luis A. Orozco is in the Department of Physics,
University of Maryland, College Park, Maryland
20742, USA.

e-mail: lorozco@physics.umd.edu

1. McKeever, J., Boca, A., Boozer, A. D., Buck, J. R. & Kimble, H. J.
Nature **425**, 268–271 (2003).

2. Raithel, G., Wagner, C., Walther, H., Narducci, L. M. & Scully, M. O. *Adv. At. Mol. Opt. Phys. Suppl.* **2**, 57–121 (1994).
3. Brune, M., Raimond, J. M., Goy, P., Davidovich, L. & Haroche, S. *Phys. Rev. Lett.* **59**, 1899–1902 (1987).
4. An, K., Childs, J. J., Dasari, R. R. & Feld, M. S. *Phys. Rev. Lett.* **73**, 3375–3378 (1994).
5. Jaynes, E. T. & Cummings, F. W. *Proc. IEEE* **51**, 89–109 (1963).
6. Paul, H. *Ann. Phys. (NY)* **11**, 411–412 (1963).
7. Mu, Y. & Savage, C. M. *Phys. Rev. A* **46**, 5944–5954 (1992).

Cancer

Cues for migration

René Bernards

Lack of oxygen causes the cells of certain tumours to spread to new locations. It also activates a homing mechanism that enables the migrating cells to target specific organs.

The ability of tumour cells to metastasize — to spread to other parts of the body — is perhaps the main reason that certain types of cancer are often fatal. But how do tumours acquire this characteristic? Starving tumour cells of oxygen seems to be one trigger for metastasis, and researchers are beginning to uncover the molecular pathways that underlie this phenomenon^{1,2}. Writing on page 307 of this issue, for instance, Staller and co-workers³ reveal that a gene called *CXCR4* is activated by the lack of oxygen, and that this activation causes tumour cells to migrate and to home in on a specific set of organs.

Highly aggressive tumours rapidly outgrow their blood supply, leaving the cells starved of oxygen — a condition known as

hypoxia. Tumour cells adapt to hypoxia by increasing their synthesis of a protein named HIF (hypoxia-inducible factor), which in turn binds to and activates several genes⁴. The proteins encoded by these HIF-responsive genes have a variety of functions. Some increase tissue oxygenation — such as vascular endothelial growth factor (VEGF), which stimulates the outgrowth of new blood vessels — and some enhance cellular glucose uptake and metabolism to allow energy generation when oxygen is scarce.

Our understanding of how levels of HIF are upregulated during hypoxia is growing rapidly⁴. The protein encoded by the von Hippel–Lindau tumour suppressor gene (pVHL), which is frequently mutated in cancer, is central to this process. The normal

VHL protein is part of a complex that, when oxygen is abundant, targets the α -subunits of HIF (HIF- α) for degradation. Recognition of these subunits by pVHL depends on a modification of HIF- α that can occur only in the presence of oxygen. When oxygen is scarce, this modification does not occur and so HIF- α escapes destruction, causing an increase in HIF levels and enhancing expression of the hypoxia-inducible genes. Mutations of the *VHL* gene produce an effect rather like hypoxia — mutant forms of pVHL found in cancer cannot destroy HIF- α , and as a result, the hypoxia-inducible genes are persistently activated⁴ (Fig. 1).

In the new study, Staller *et al.*³ searched for genes that are regulated by pVHL. They introduced *VHL* into renal carcinoma cells (which lack a normal copy of this gene) and then, using DNA microarray analysis, they looked for changes in the activity of thousands of other genes, under non-hypoxic conditions. Unexpectedly, they found that normal pVHL dramatically reduced the production of a receptor protein called CXCR4. This receptor binds chemokines — secreted proteins, rather like growth factors, that allow migrating cells (immune cells, for example) to navigate to specific organs⁵. The binding of chemokines to receptors such as CXCR4 on the surface of migrating cells stimulates both cell adhesion and motility, and causes the cells to move towards the source of the chemokine.

But how does pVHL regulate the production of CXCR4? As the authors expected, it does so by downregulating HIF. The authors found a functional HIF-binding site in the regulatory region of the *CXCR4* gene. They also found that cells containing normal pVHL produced more CXCR4 when exposed to hypoxia. So *CXCR4* seems to be a bona fide hypoxia-inducible gene.

The connection between CXCR4 and hypoxia is revealing, because several studies have indicated that 'chemoattraction' through CXCR4 contributes to organ-specific metastasis in certain forms of cancer. For instance, human breast cancer cells often contain high levels of CXCR4, and these cells preferentially metastasize to sites that produce large amounts of the chemokine SDF-1 α (the binding molecule, or ligand, of the CXCR4 receptor), such as the lungs and bone marrow⁶. In fact, *CXCR4* is part of a small set of genes that cooperate to promote bone metastases from breast cancer⁷.

But why does hypoxia induce CXCR4? Presumably, it doesn't matter to a tumour cell where it migrates. An answer to this question can be deduced from the characteristics of mice that are genetically engineered to lack this gene. Such mice show defects in the branching and/or remodelling of certain blood vessels⁸. So the activation of CXCR4 in blood-vessel cells might be part of an integrated hypoxic response that allows

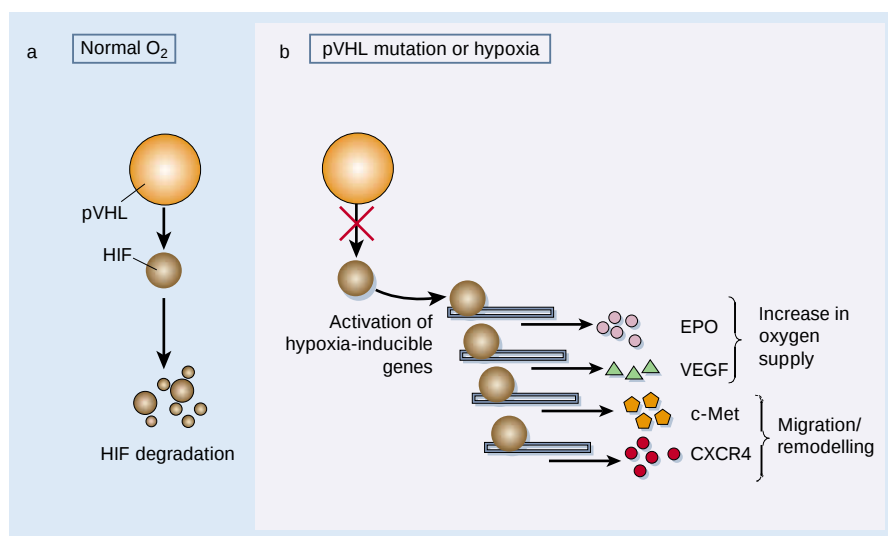


Figure 1 The hypoxic response. a, Under conditions of normal oxygen, the von Hippel–Lindau tumour suppressor protein (pVHL) modifies the protein HIF, which leads to its destruction. b, When oxygen is scarce (hypoxia), or when pVHL is mutated, HIF accumulates inside cells and activates the expression of certain genes. This triggers two complementary responses. First, tissue oxygenation is stimulated through the activation of genes such as VEGF (which stimulates the outgrowth of new blood vessels) and erythropoietin (EPO, which stimulates the production of red blood cells). Second, tumour cells are stimulated to move away from the site of hypoxia through the activation of genes such as *c-Met*, which enhance cell motility and invasion. Now, Staller *et al.*³ show that the gene *CXCR4* is also activated by HIF. CXCR4 not only stimulates migration, it also enables tumour cells to home in on specific, distant organs.

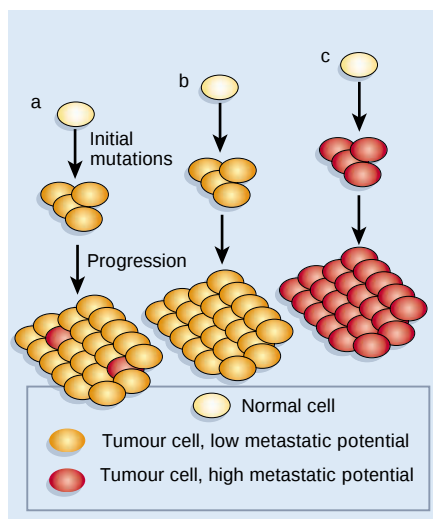


Figure 2 The development of metastatic capability in cancer cells. a, The classical view is that tumour cells with a high metastatic capability arise at low frequency within the primary tumour. These rare variants give rise to metastases. b, c, DNA-microarray-based gene-expression profiling has indicated, however, that some tumours are homogeneously of 'good prognosis' and others are uniformly of 'poor prognosis'. One way of explaining these data is to assume that some tumours start off on the wrong foot, their initial mutations predisposing them to become highly metastatic after additional mutations have been acquired^{12,13}. Other incipient tumours have initial mutations that make progression to a metastatic phenotype unlikely. The work of Staller *et al.*³ provides support for the situation indicated in c. Mutations in the *VHL* gene may not only stimulate uncontrolled cell division (an early step in cancer development), but also, as the authors show, promote cell motility, invasion and homing to distant organs.

both the generation of new vessels (through the induction of VEGF) and the remodelling of existing vessels (through CXCR4 induction). The finding that VEGF induces expression of CXCR4 supports the view that CXCR4 plays a part in remodelling the vasculature during hypoxia⁹.

CXCR4 might be needed to promote the survival of tumour cells in a hypoxic environment¹⁰, and it enhances cell motility, allowing the tumour cell to migrate away from areas of low oxygen. Hypoxia-induced metastasis also occurs through other signalling pathways — for instance, the gene encoding the c-Met receptor was recently identified as being hypoxia-inducible¹. This receptor protein enhances both cell motility and invasion through binding its ligand, hepatocyte growth factor. But the CXCR4 pathway is different because it enables the migrating cells to navigate to specific organs. So, apart from the well-known effects of hypoxia on blood-vessel growth, it also seems to trigger a second and complemen-

tary response, enabling cells to migrate away from areas of low oxygen and to home to specific, distant organs (Fig. 1). That CXCR4 expression stimulates homing to distant sites is probably not relevant to the physiological response to hypoxia, but only an unfortunate side effect of tumour-cell hypoxia. This side effect could, however, explain the generally worse prognosis of patients with a hypoxic tumour². Consistent with this, Staller *et al.* show that clear-cell renal cancers that harbour a mutant form of *VHL* express high levels of CXCR4, which correlates with poor survival³.

The new study also contributes to the continuing debate over whether the ability of tumour cells to metastasize is acquired early or late^{11,12} (Fig. 2). The data of Staller *et al.* enable us to envisage how tumour cells might be poised early on to spread to other parts of the body. Incipient tumour cells that acquire a mutation in the *VHL* gene early on may be predestined to spread to secondary sites at a later stage — when they have acquired additional growth-promoting

mutations — as the activation of genes such as *c-Met* and *CXCR4* endows them with the ability to migrate, invade and home in on specific tissues. So as well as providing clues about how tumours metastasize, Staller and colleagues' findings edge us closer to an understanding of the timing of tumour progression.

René Bernards is in the Division of Molecular Carcinogenesis and Center for Biomedical Genetics, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
e-mail: r.bernards@nki.nl

1. Pennacchietti, S. *et al.* *Cancer Cell* **3**, 347–361 (2003).
2. Bottaro, D. P. & Liotta, L. A. *Nature* **423**, 593–595 (2003).
3. Staller, P. *et al.* *Nature* **425**, 307–311 (2003).
4. Kaelin, W. G. Jr *Nature Rev. Cancer* **2**, 673–682 (2002).
5. Homey, B., Muller, A. & Zlotnik, A. *Nature Rev. Immunol.* **2**, 175–184 (2002).
6. Muller, A. *et al.* *Nature* **410**, 50–56 (2001).
7. Kang, Y. *et al.* *Cancer Cell* **3**, 537–549 (2003).
8. Tachibana, K. *et al.* *Nature* **393**, 591–594 (1998).
9. Salcedo, R. *et al.* *Am. J. Pathol.* **154**, 1125–1135 (1999).
10. Zeelenberg, I. S., Ruuls-Van Stalle, L. & Roos, E. *Cancer Res.* **63**, 3833–3839 (2003).
11. Bernards, R. & Weinberg, R. A. *Nature* **418**, 823 (2002).
12. Hynes, R. O. *Cell* **113**, 821–823 (2003).
13. Bernards, R. & Weinberg, R. A. *Nature* **419**, 560 (2002).

Palaeobotany

Fishing for the first plants

Paul Kenrick

Sifting of organic residues from ancient rocks has netted a catch of tiny fossils that provide clues about when plant life first appeared on land.

The significance of microscopic spores entombed in rocks dating from about 440–470 million years ago has puzzled investigators of early life. Are these spores proof that plant life existed on land long before the time suggested by other forms of fossil evidence? And, if so, what sorts of plants do they represent? On page 282 of this issue, Wellman *et al.*¹ present direct evidence of the life forms that produced these enigmatic spores, and their findings lend credence to the notion that minute plants existed on land 470 million years ago.

Spores are produced by land plants in prodigious quantities. These robust, decay-resistant particles can become incorporated into sediments, providing a record of floral change through geological time. Careful study of rocks dating from the Ordovician period, 443–489 million years ago, has revealed an unexpected diversity of spores that are much older than the fossilized remains of plants that could have produced them^{2,3}. These minute grains might represent evidence of the earliest land flora, but how can we be sure that these spores came from bona fide land plants, and what can they tell us about the nature of this early flora?

Answering these questions is difficult

because the most ancient spores are rather odd. Instead of being dispersed as single grains, many are fused in pairs or in groups of four, and some are enclosed in an extra membrane^{2,3} (Fig. 1). These so-called permanent diad and permanent tetrad configurations are unlike the spores of most living plant species, but they do bear some resemblance to the spores of certain present-day liverworts. Diad and tetrad spores have also been found in land-plant fossils dating from an early part of the Devonian period (400–417 million years ago)^{4,5}.

So, one school of thought holds that the tetrads and diads of the Ordovician period are evidence of land plants that are related to living bryophytes (liverworts, mosses and their kin)^{2–5}. Others contend, however, that the data linking these spores to bryophytes are too tenuous⁶. The spore-producers might be close relatives of land plants, but that does not necessarily make them bryophytes. It is conceivable that in both ecological and physiological terms they were little more than aquatic algae. Direct evidence of the life forms that produced the Ordovician spores could settle this matter, but so far this has proved elusive.

Wellman *et al.*¹ take us a step closer to resolving this controversy. They used standard

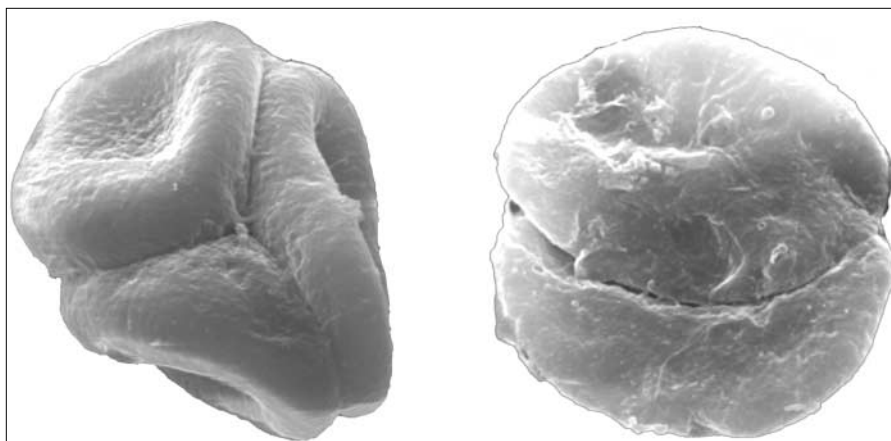


Figure 1 What the first land plants left behind. Evidence of the earliest land plants comes from spores extracted from 440–470-million-year-old rocks in Oman. The nature of the plants that produced these grains has remained a mystery, but Wellman *et al.*¹ have now found evidence that the spores were produced by minute land-dwelling plants that may have resembled liverworts. Unlike most of their modern counterparts, many of the earliest spores were dispersed in groups of four (tetrads; left) or in groups of two (diads; right). Spore diameter is 20 μm .

spore-extraction methods to recover organic residues from Ordovician sediments collected from a borehole in Oman. When they passed the insoluble organic material through a series of sieves designed to trap plant fragments and spores of various sizes, the authors found many well-preserved spores — but they also found something much more intriguing. Trapped in the sieve containing the largest pores were elongated, disc-shaped objects that, on closer inspection, proved to be clumps of spores packaged in a type of cuticle. These fossil fragments are exciting because they resemble the spore-bearing organs of later land plants. So Wellman and colleagues' trawl through organic residues had netted a catch of the tiny plants that produced the spores. Although far from complete, these specimens indicate that the Ordovician spores were indeed produced by land plants and not by algae. But what did these early plants look like, and how are they related to modern plant forms?

The plants that produced the spores were certainly minute and probably simple, but these frustratingly incomplete fragments tell us little else. In a hunt for further clues, Wellman and colleagues looked in detail at the structure of the spore wall. They found a laminate interior resembling that of some living liverworts, which is consistent with other evidence^{7,8} pointing to liverworts as the closest living relatives of the earliest land plants. Although intriguing, the spore-wall data are unlikely to win over the sceptics⁹. What we need now are more complete specimens of these tiny plants, and Wellman and colleagues' recovery techniques are clearly promising ways to achieve this.

One further issue raised by the new study is the possibility that plant life existed on land even before the Ordovician period. Could signs of plant life be awaiting discovery in even older rocks? Maybe so, according

to one study⁹. These authors investigated the time of origin of land plants using a 'molecular clock' analysis, in which the timing of evolutionary divergence is inferred from comparisons of the gene sequences of living plant species. Controversially, this approach placed the common ancestor of all living land plants in the Precambrian period, around 700 million years ago. From a

palaeontological perspective, the existence of land plants this long ago seems very unlikely. For one thing, there is no unequivocal evidence for land-plant spores in rocks predating the Ordovician period. Given the ubiquity of such spores in near-shore marine sediments throughout later periods, this dearth of microfossil data is telling. And different clocks tell different times — a more recent molecular analysis placed the origin of land plants close to the time predicted from the spore findings¹⁰. If plant life on land did predate the Ordovician period then it is very well hidden indeed. ■

Paul Kenrick is in the Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.

e-mail: P.Kenrick@nhm.ac.uk

1. Wellman, C. H., Osterloff, P. L. & Mohiuddin, U. *Nature* **425**, 282–285 (2003).
2. Edwards, D. & Wellman, C. H. in *Plants Invade the Land: Evolutionary and Environmental Perspectives* (eds Gensel, P. G. & Edwards, D.) 3–28 (Columbia Univ. Press, New York, 2001).
3. Gray, J. *Phil. Trans. R. Soc. Lond. B* **309**, 167–195 (1985).
4. Edwards, D., Duckett, J. G. & Richardson, J. B. *Nature* **374**, 635–636 (1995).
5. Wellman, C. H., Edwards, D. & Axe, L. *Bot. J. Linn. Soc.* **127**, 117–147 (1998).
6. Chaloner, W. G. in *Proc. XIV Int. Bot. Congr.* (eds Greuter, W. & Zimmer, B.) 301–316 (Koeltz, Berlin, 1988).
7. Nickrent, D. L., Parkinson, C. L., Palmer, J. D. & Duff, R. J. *Mol. Biol. Evol.* **17**, 1885–1895 (2000).
8. Qiu, Y. L. & Lee, J. J. *Phycol.* **36**, 799–802 (2000).
9. Heckman, D. S. *et al. Science* **293**, 1129–1133 (2001).
10. Sanderson, M. J. *Am. J. Bot.* **90**, 954–956 (2003).

Global change

Probing early atmospheres

Stephen J. Mojzsis

Information about atmospheric conditions far back in Earth's history is embedded in the isotopic composition of fossil microbes. Such studies are technically demanding, but hold considerable promise.

What were atmospheric CO₂ concentrations like during the Proterozoic eon, the interval of Earth's history between about 2,500 million and 543 million years ago? The issue is germane to studies of the ancient biosphere because, owing to the way in which the Sun's composition and other properties evolved, its luminosity is thought to have increased gradually over time. Aside from a small contribution from internal heating that is driven by radioactive decay, the surface temperatures of a planet are governed by the amount of solar radiation it receives and how this interacts with the atmosphere. The fact that greenhouse gases, chiefly CO₂ (but also water vapour and methane), re-radiate infrared radiation to the surface keeps the present average surface temperature some 33 K above what it would otherwise be. Models of solar luminosity¹ indicate that if the greenhouse had not been much greater early in Earth's history, global

glaciation should have held sway throughout the Proterozoic and the preceding Archaean eon, which stretches back to about 3,900 million years ago.

However, various lines of reasoning suggest that, over geological time, Earth has maintained surface pressures and temperatures that were conducive to the presence of liquid water, and thereby a global habitat for life. Carbon dioxide has always been an important greenhouse gas. But how much was present in the atmosphere during the Proterozoic, when the surface biosphere became transformed from a wholly microbial and anoxic world to an oxygenated domain poised for the emergence of diverse multicellular life?

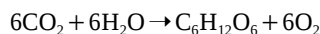
On page 279 of this issue², Kaufman and Xiao describe their use of a novel approach to this question in comparing the carbon-isotope compositions of eukaryotic algal microfossils (acritarchs) to those of 'con-sanguineous carbonate' — that is, carbonate



Figure 1 Shot on site — an outcrop of the Ruyang Group in Shanxi Province, China.

produced from CO_2 by inorganic processes at around the same time. They interpret well-characterized acritarchs, extracted from shales of the Ruyang Group in North China (Fig. 1), as representing the remains of marine organic matter produced during active photosynthesis³. The ages of the Ruyang Group rocks are only imprecisely known. But from correlation with the Roper Belt sediments of Western Australia, which contain similar fossil forms⁴ and have recently been investigated for sulphur-

isotope variability⁵, they are estimated to be around 1,400 million years old — that is, mid-Proterozoic in age. Microbes are important players in long-term biogeochemical cycles, not least by maintaining an oxygen-rich atmosphere through the oxygenic photosynthetic reaction:



Because a short-circuit exists in the marine carbon cycle, whereby some organic carbon

($\text{C}_6\text{H}_{12}\text{O}_6$) is stored as sedimentary organic matter, a net leakage of oxygen to the atmosphere results. Carbonate carbon and organic molecules are respectively derived from CO_2 through precipitation in water and photosynthesis. Isotopic effects associated with these reactions lead to a fractionation of the $^{13}\text{C}/^{12}\text{C}$ ratio, between organic carbon and carbonate carbon that is in equilibrium with atmospheric CO_2 , of about 25 parts per thousand (25‰). So organic materials are depleted in ^{13}C relative to consanguineous carbonate.

At present, marine photosynthesis responsible for about 99% of primary productivity is performed by single-celled eukaryotic algae using the Calvin cycle⁶ — a major pathway for oxygenic photosynthesis. Kaufman and Xiao's approach² makes sense, because the source carbon in photosynthesis is CO_2 , and the morphologically complex organisms found in the Ruyang shales are reasonably interpreted as eukaryotic organisms that used the Calvin cycle. As a result, it becomes possible to estimate the CO_2 concentration in the system using the magnitude of carbon-isotope fractionation (ϵ_p) between organic carbon and carbonate. This tactic works if the carbon-isotopic composition of individual acritarchs can be measured.

For this purpose, Kaufman and Xiao used an ion-microprobe technique, in which a beam of caesium ions is focused on minute carbonaceous particles to measure their $^{13}\text{C}/^{12}\text{C}$ ratio at high mass resolution. This is an improvement over attempts to measure carbon isotopes in individual microfossils⁷. Kaufman and Xiao isolated the acritarchs from the Ruyang shales by acid extraction, allowing for rapid repeat carbon-isotope measurements of separate fossils and ensuring

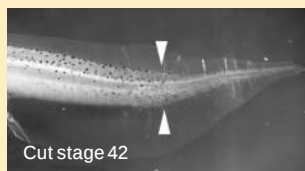
Developmental biology

A tadpole's tale

If tadpoles of the African clawed toad *Xenopus laevis* lose their tail, they can usually make another — but not always. So say Caroline W. Beck and colleagues, who have identified a 'refractory period' during which tails cannot regenerate (*Dev. Cell* **5**, 429–439; 2003). Studies of this down-time have provided insight into the molecular mechanisms that underlie tail regeneration.

The ability to regenerate limbs is scattered throughout the animal kingdom, even occurring in the adults of some species — certain lizards, for instance. But the phenomenon is not well understood at the molecular level. Beck *et al.* have used a variety of new molecular techniques to tackle this question.

The authors started by



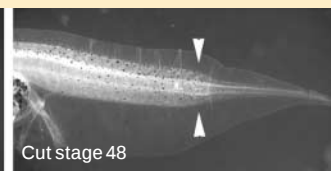
investigating tail regeneration at different stages of tadpole development. They found that 3 and 7 days into development, at stages 42 and 48, around two-thirds of the tadpoles could regrow their tails after amputation (see picture; the arrowheads represent the position at which the tail was amputated). At stage 49 and beyond, that figure was near 100%. But at stages 46 and 47, only a very few of the tails regenerated (although the tadpoles developed into froglets normally).

What's going on during this



refractory period? A look at the regenerating tails showed that a thin skin formed over the cut stump, and that unspecialized cells accumulated beneath the skin; these later began to produce the spinal cord, muscle cells and other cells required for tail formation. This did not occur in the non-regenerating tails.

Proteins involved in the bone morphogenetic protein (BMP) and Notch signalling pathways are required for embryonic tail development. Might they also be expressed in regenerating tails? Beck



et al. show that they are, but that they are not expressed in tails amputated during the refractory period. If, however, the BMP receptor is forcibly expressed during this period, tail regeneration occurs in nearly all cases. Further findings hint that the BMP pathway activates the Notch pathway to regenerate the spinal cord, but works independently of Notch to regenerate muscle. Whether these molecular events underlie the remarkable phenomenon of regeneration more generally remains to be seen. **Amanda Tromans**

sample homogeneity, so it can be argued that no significant amount of secondary organic matter has affected the isotopic composition of the primary photosynthetic product.

Analyses of the $^{13}\text{C}/^{12}\text{C}$ of individual acritarchs, coupled with calculated magnitudes of ϵ_p depending on the CO_2 levels of the growth medium and volume/surface-ratio estimates of the original algal forms, suggest that, around 1,400 million years ago, CO_2 levels were at least 10 times and possibly more than 200 times higher than at present. The results agree broadly with studies⁸ on the iron mineralogy of 'weathering profiles' (palaeosols), estimated at 2,750 million years old, which placed a minimum value for CO_2 of about 100 times those of today. But is this enough CO_2 to have kept the Proterozoic in a suitable greenhouse state?

Models of solar evolution⁹ indicate that the Sun was only about 88% of its present luminosity at the time of deposition of the Ruyang Group. A typical solution to why there was no planetary freeze throughout the Archaean and Proterozoic is to invoke the existence of CO_2 levels on the early Earth of around 1,000 times present levels¹⁰. Such a CO_2 -rich atmosphere can be qualitatively justified by theoretical feedback mechanisms that link increased volcanism and decreased continental weathering resulting in mineral dissolution in liquid water, and thereby to higher atmospheric CO_2 concentrations.

The new results² appear to be close to the limits required in atmospheric models to overcome diminished solar luminosity¹¹. Other workers have proposed that a combination of gases, probably including a large contribution from methane produced by living organisms, kept the early Earth from being permanently frozen over¹². But once free oxygen became a major atmospheric constituent between 2,300 million and 2,100 million years ago, it shortened the

photochemical lifetime of methane, diminishing its greenhouse influence. All in all, it would seem that the imposition of high, steady-state concentrations of CO_2 is consistent with the temperature record of a mostly ice-free Proterozoic.

Studies of the early Earth are dogged by the meagre evidence available: discontinuities punctuate the entire sweep of the ancient rock record. But they can be tackled by intertwining strands of evidence from, for instance, geochemistry, geophysics, 'fossil morphometrics' and analyses of the molecular phylogenetic relationships between living organisms. 'Geobiology' is an innovative and inspiring venture in the Earth sciences¹³, and by linking metabolism to CO_2 concentrations, Kaufman and Xiao have provided a promising way of improving our understanding. The way forward is to pursue independent proxy measures of atmospheric CO_2 levels, and the isotopic payload of fossil microbes may prove an excellent source of such measures. ■

Stephen J. Mojzsis is in the Department of Geological Sciences, Center for Astrobiology, University of Colorado, Boulder, Colorado 80309-0399, USA.

e-mail: mojzsis@colorado.edu

1. Sagan, C. & Mullen, G. *Science* **177**, 52–56 (1972).
2. Kaufman, A. J. & Xiao, S. *Nature* **425**, 279–282 (2003).
3. Xiao, S., Knoll, A. H., Kaufman, A. J., Yin, L. & Zhang, Y. *Precamb. Res.* **84**, 197–220 (1997).
4. Javaux, E. J., Knoll, A. H. & Walter, M. R. *Nature* **412**, 66–69 (2001).
5. Shen, Y., Knoll, A. H. & Walter, M. R. *Nature* **423**, 632–635 (2003).
6. Tyrrell, T. *Nature* **400**, 525–531 (1999).
7. House, C. H. *et al.* *Geology* **28**, 707–710 (2000).
8. Rye, R., Kuo, P. H. & Holland, H. D. *Nature* **378**, 603–605 (1995).
9. Bahcall, J. N., Pinsonneault, M. H. & Basu, S. *Astrophys. J.* **555**, 990–1012 (2001).
10. Walker, J. C. G., Hays, P. B. & Kasting, J. F. *J. Geophys. Res.* **86**, 9776–9782 (1981).
11. Kasting, J. F. *Science* **259**, 920–926 (1993).
12. Pavlov, A. A., Hurtgen, M. T., Kasting, J. F. & Arthur, M. A. *Geology* **31**, 87–90 (2003).
13. Knoll, A. H. *Geobiology* **1**, 3–14 (2003).

Viral genetics

Deadly partnerships

Steven A. Frank

Pairs of viral genomes work together to destroy their hosts more quickly. How this might occur remains unknown, but study of the phenomenon should provide insight into how genetic systems evolve.

Early in the history of life, different copies of replicating nucleic acids must have existed near each other. Some of these genomes probably parasitized their neighbours by becoming shorter, dropping essential information and using proteins encoded by the full-length molecules. The shorter parasitic genomes might have replicated faster and out-competed their fully endowed neighbours. Other pairs probably complemented each other to mutual benefit,

favouring some method for the pair to disperse together. Viral systems provide our best window back through time, allowing us to glimpse how multi-copy genetic systems might have evolved. There are many known examples of shortened viral genomes exploiting functional partners¹, but writing in the *Proceedings of the Royal Society*, López-Ferber and colleagues² now show that defective viral genomes are not always parasitic. They provide evidence that shortened



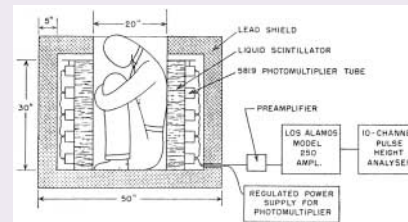
100 YEARS AGO

From the study of rays of measurable wavelengths we have lately sailed under the guidance of M. Henri Becquerel into another region where it is doubtful whether all the rays conform to the undulatory theory. In fact some of the rays are believed to be charged particles of matter, charged, that is to say, with electricity. Beyond doubt they are possessed of very extraordinary properties, inasmuch as they are able to penetrate the clothing, celluloid, gutta percha, glass, and various metals. They are, moreover, endowed with a no less remarkable physiological action, producing blisters and ulcerations in the flesh which are difficult to heal... From this we can quite understand that there is no exaggeration in the statement attributed to the discoverer, Prof. Curie, ... that he would not care to trust himself in a room with a kilogram of pure radium, because it would doubtless destroy his eyesight, burn all the skin off his body, and probably kill him.

From *Nature* 17 September 1903.

50 YEARS AGO

In the course of developing equipment for other problems, we have made some measurements of the total radioactivity content of several humans and a dog, using a technique which may have other applications in biophysics... (See Fig. 1). A dog of approximately 35 lb. weight was anaesthetized and counted in the small insert. A solution containing 0.1 μC . radium



in equilibrium with its decay products was injected into the femoral vein, and (five min. after injection) the dog was again 'counted'... The large insert was used for the measurements on humans, who were able by doubling up to be entirely within the insert... In the absence of the radium group, the potassium content of the body can be measured with good accuracy, and it is quite conceivable that application of these techniques could yield important results in the study of the role of potassium in the metabolic process. F. Reines, C. L. Cowan *et al.* From *Nature* 19 September 1953.

genomes can work with full-length partners to mutual benefit.

López-Ferber *et al.* investigated how efficiently mixed-genome infections by an insect virus called nucleopolyhedrovirus killed its larval host — in this case, the fall armyworm. First, the authors measured the dose of pure, full-length viral genomes required to kill the larvae. Host mortality is a reasonable measure of viral success because it enables transmission to new hosts — after nucleopolyhedrovirus has replicated to high densities inside the larva, it switches on genes that liquefy the insect and release the viruses to the environment.

After obtaining a baseline measure of viral success for pure, full-length nucleopolyhedrovirus genomes, the authors measured the success of mixed populations of full-length and shortened genomes. They obtained these mixtures in two ways. First, they extracted viral DNA from natural populations of virus, in which about 20–30% of the genomes contain deletions. Surprisingly, they found that this mixed viral DNA was 2.55 times as effective at killing armyworm larvae as were pure full-length genomes. In other words, it took 2.55 times the infectious dose of full-length genomes to produce the same mortality as the mixed population.

To test the pathogenicity of mixed populations in a second, more controlled way, López-Ferber *et al.* investigated a common type of shortened genome found in natural populations. This deletion genotype could not successfully infect the host by itself, but it could succeed when in a mixed infection. The authors created five mixtures containing 10%, 25%, 50%, 75% and 90% of the shortened genome combined with the full-length genome. The most pathogenic combination contained 25% of the shortened genome and it was 2.91-fold as effective at killing as pure full-length genomes. So the maximum pathogenicity in controlled experiments occurred with about the same ratio of short to full-length genomes as is found in natural nucleopolyhedrovirus populations.

Surprisingly, these experiments demonstrate a mutually beneficial interaction between short and full-length genomes. López-Ferber *et al.* do not provide a mechanistic explanation for this phenomenon, but it is worth considering a few hypotheses because the real interest in this story is what can be learned in general about viral population genetics. To understand how this particular case might lead to a deeper insight, we must first look more closely at the biological details³.

The nucleopolyhedrovirus has an interesting method of packaging its genomes for transmission between hosts (Fig. 1). Unusually, several viral genomes are packaged together in membranous envelopes derived from larval cells. The multi-genome virus envelopes embed themselves in larger,

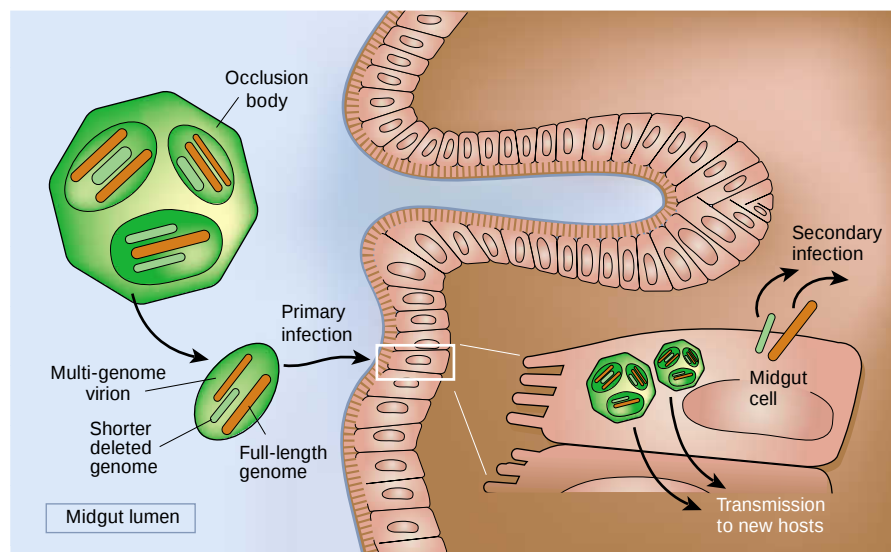


Figure 1 The nucleopolyhedrovirus life cycle. Infection begins when a larva eats a proteinaceous occlusion body, containing many virus envelopes. Each envelope contains a mixture of viral genomes, some of which can contain deletions. The occlusion body dissolves in the larval gut, releasing the enveloped viruses, which establish a primary infection in the midgut cells. The infected midgut cell first produces single viral genomes that bud from the cell surface and infect other cells — it is estimated that around four virus genomes invade each cell during this secondary infection⁶. As viral multiplication continues within the infected cell, occlusion bodies form. After many host cells have been infected, the virus turns on genes that liquefy the host, releasing occlusion bodies, which are eaten by new hosts.

proteinaceous structures called occlusion bodies, which are released when the larva is liquefied. When other larvae ingest the occlusion bodies, they dissolve in the gut to release the virus envelopes and the infection cycle begins again.

The particular shortened genotype investigated by López-Ferber *et al.* lacked a gene called *pif*. The protein, PIF, encoded by this gene enables the virus envelopes to invade the midgut cells — the site of primary infection^{2,4}. If envelopes containing shortened genomes are injected directly into a larva, infection proceeds normally but the new virus envelopes produced do not contain PIF. As a result they cannot invade midgut cells in new larvae to start another cycle of infection.

Lack of the *pif* gene explains why a shortened genome needs a full-length partner, but it does not explain why the full-length genome does better with a shortened partner. One hypothesis is that paired genomes can complement defects in partners⁵ — certain genes carried by full-length genomes might harbour potentially deleterious mutations that are masked by the normal code carried by its partner. Such complementation is not restricted to combinations of full-length and shortened viral genomes: it is a general explanation for why multiple copies of genomes are often packaged together. Complementation must have arisen with the earliest genetic systems and it continues today in many organisms — even our own genomes have two copies.

A second hypothesis to explain how long

and short genomes might interact to mutual benefit is that the shortened genome encodes a useful trait not carried by the full-length genome. Because the shortened genome must always partner a full genome that has the normal code, the short genome is free to accumulate mutations that lead to new benefits. For example, it might evolve a function that enables mixed infections to kill alternative host species more effectively.

Exactly how the example of mutualism described by López-Ferber *et al.*² arose remains to be seen. But their results suggest that insect viruses such as nucleopolyhedrovirus could be effective, general models to study how complementation might promote the packaging together of several viral genomes and how these genomes might then diverge to take on mutually beneficial traits. So, beyond destroying their hosts more efficiently, mixed nucleopolyhedrovirus infections might teach us how genetic systems evolve.

Steven A. Frank is in the Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, USA.

e-mail: safrank@uci.edu

- Knipe, D. M. & Howley, P. M. *Fields Virology* 4th edn (Lippincott, Williams & Williams, Philadelphia, 2001).
- López-Ferber, M., Simón, O., Williams, T. & Caballero, P. *Proc. R. Soc. Lond. B* doi:10.1098/rspb.2003.2498 (2003).
- Volkman, L. E. *Adv. Virus Res.* **48**, 313–348 (1997).
- Kikhno, I., Gutierrez, S., Croizier, L., Croizier, G. & López-Ferber, M. *J. Gen. Virol.* **83**, 3013–3022 (2002).
- Rohrman, G. F. in *The Biology of Baculoviruses* (eds Granados, R. R. & Federici, B. A.) 203–215 (CRC, Boca Raton, Florida, 1986).
- Bull, J. C., Godfrey, H. C. J. & O'Reilly, D. R. *Appl. Environ. Microbiol.* **67**, 5204–5209 (2001).

Molecular biology

Telomere stutters
superfluous

Genes Dev. 17, 2271–2282 (2003)

Telomeric repeats — stretches of recurring DNA sequences — are an integral part of the telomeres that cap the ends of chromosomes. But Mahito Sadaie *et al.* have shown that chromosomes can work pretty well without them.

Researchers have found before that fission yeast survive when their telomeric repeats are deleted by fusing the yeast's linear chromosomes into circles. Now Sadaie *et al.* show that one key function of telomeres is also maintained. Telomeres are important during meiosis — the type of cell division that generates reproductive cells — in that they cluster together, thereby allowing chromosome pairing. But Sadaie *et al.* find that the repeat-less circular chromosomes can still pair up. Moreover, Taz1, a protein that binds telomeric repeats, remains latched onto the repeat-less chromosomes. The group suggests that a genetic sequence in a region called the subtelomere serves as a back-up mechanism to recruit such telomeric proteins.

But telomeric repeats are needed for something. Cells lacking repeats cannot rebuild fully working telomeres after they are dismantled, as the authors showed by creating mutants lacking Taz1 and then resupplying the protein. This shows that the three-dimensional structure of the repeats may provide a scaffold for Taz1 and other proteins.

Helen Pearson

Atomic physics

Squishy ball lightning

Appl. Phys. Lett. 83, 2283–2284 (2003)

Ball lightning poses an irresistible puzzle. Nikola Tesla and Peter Kapitsa are among those who have previously sought to find explanations for these glowing balls, typically about the size of a child's head, that float through the air during thunderstorms and sometimes pass through closed doors and windows. They are widely thought to be balls of plasma, and might be akin to the luminous plasmas occasionally formed in the detonation of explosives.

One of the most striking and unexplained properties of these ghostly apparitions is their cohesion. If confined in glass tubing, explosion plasmas can be guided around sharp corners, whereas ball lightning can squeeze through narrow openings. J. J. Gilman suggests that this cohesion might arise in a plasma of Rydberg atoms, which are atoms with an outer (valence) electron that has been excited into an orbital with an extremely high quantum number (10^3 – 10^4). The radius of such an orbital may be

macroscopic: up to several centimetres. This makes the electron cloud very 'floppy', so that the atom is highly polarizable and experiences strong van der Waals attractive interactions. Gilman estimates that a plasma containing such Rydberg atoms would be a squishy ball with about one-hundredth the cohesive energy per atom of a metal. Philip Ball

Molecular biology

Spreading like wildfire

Science 301, 1545–1547 (2003)

Gene expression in nematode worms can be shut down by a process called RNA interference (RNAi). An injection of double-stranded RNA (dsRNA) that matches a particular gene suffices to silence that gene throughout the worm — and in its progeny. Evan H. Feinberg and Craig P. Hunter now suggest how a protein called SID-1 may mediate this ripple effect.

The natural function of RNAi in worms is not certain, but in some other organisms it protects against viral diseases. Viruses often carry their genetic information on a single strand of RNA, but make double-stranded copies when infecting their host. This triggers the host's defences, which somehow stop the RNA from being translated into functional proteins. In worms, such viruses have not yet been identified; nevertheless, researchers exploit RNAi to study gene function. By introducing specific dsRNA sequences, a worm can be seduced to slice up its own RNA and thus silence gene expression.

Previously, Hunter and colleagues showed that the gene *sid-1* is essential for spreading the gene-silencing effects of RNAi throughout a worm. Now they find that the SID-1 protein allows cells to soak up dsRNA. Although it is possible that SID-1 facilitates dsRNA uptake indirectly, the authors propose that it forms a channel that allows the intercellular transport of dsRNA by diffusion.

Marie-Thérèse Heemels

Behavioural ecology

Horns and lifestyle

Behav. Ecol. Sociobiol. doi:10.1007/s00265-003-0672-6 (2003)

Animals display an impressive array of horns and antlers, from the smooth twisted horns of the kudu, to the curved battering-rams of bighorn sheep, to the many-pointed 'racks' of moose (pictured). But behind this diversity lie some common ecological patterns, T. M. Caro and colleagues have discovered.

The researchers carried out a comparative phylogenetic survey of the headgear and habits of 39 species of deer and 124 species of bovid (which include cattle, antelope, sheep and goats). They find that an animal's



K. WARD/CORBIS

weapons tend to match its social life. Monogamous and solitary species tend to have straight horns whose tips face in, whereas monogamous deer have simple antlers bearing two to five points.

Polygamous, group-living species have a wider range of armaments, which are more likely to be twisted and outward facing. Caro *et al.* suggest that this reflects the greater competition in social species — between males for mates and females for food — and so the style and intensity of fighting. In contrast, it seems that environment has little influence on horn shape. For example, forest dwellers can apparently possess complex weapons even though they have to dodge through dense vegetation.

John Whitfield

Meteorology

Oscillating eyes

Geophys. Res. Lett. doi:10.1029/2003GL017653 (2003)

Chun-Chieh Wu and colleagues have peered inside Typhoon Zeb, which hit Luzon in the Philippines in 1998, tracking the evolution of its eye. The eye squinted before the storm made landfall, then closed altogether before reappearing wider than ever as the cyclone whirled out to sea — a course of events that the authors have been able to model by computer.

The eye of a tropical cyclone, the calm, windless centre, is surrounded by the eyewall. This is the most intense region, with maximum wind speeds and rainfall, and predicting a storm's course and intensity depends greatly on understanding the eyewall.

Wu *et al.* studied satellite data of Zeb's behaviour, then modelled it by numerical simulation. As well as simulating the eye shrinking to 50 km in diameter over the coast of Luzon, they were able to chart its growth to 200 km with an incomplete eyewall as it headed out to sea, as well as its subsequent narrowing again as the typhoon approached Taiwan.

The aim of this line of work is to improve knowledge of how factors such as ocean temperature and surface terrain influence the development of tropical cyclones, and so help to create better forecasting models.

Tom Clarke

Delivery of pollutants by spawning salmon

Fish dump toxic industrial compounds in Alaskan lakes on their return from the ocean.

Pollutants are widely distributed by the atmosphere and the oceans¹. Contaminants can also be transported by salmon and amplified through the food chain. Here we show that groups of migrating sockeye salmon (*Oncorhynchus nerka*) can act as bulk-transport vectors of persistent industrial pollutants known as polychlorinated biphenyls (PCBs), which they assimilate from the ocean and then convey over vast distances back to their natal spawning lakes. After spawning, the fish die in their thousands — delivering their toxic

cargo to the lake sediment and increasing its PCB content by more than sevenfold when the density of returning salmon is high.

Sockeye salmon spawn in freshwater after spending most of their life in the ocean, and die after spawning. They acquire more than 95% of their biomass in the north Pacific Ocean^{2,3} and then return to their nursery lakes, which may be more than 1,000 km from the ocean⁴. The lipids accumulated by salmon to fuel this long upstream migration sequester PCB compounds: from a PCB concentration of less than 1 ng per litre in the ocean^{5,6}, sockeye salmon accumulate PCBs to a concentration of 2,500 ng per gram of lipid⁷. This means that one million adult salmon could transport more than 0.16 kg of PCBs to the spawning area, which is comparable to the amount of fugitive PCBs released annually from hazardous waste incinerators⁸. Fat-soluble PCBs are transported to freshwaters, where they enter aquatic and terrestrial foodwebs⁷.

During 1995, 1997, 1998 and 2002, we collected sediment cores from eight lakes, providing a range of salmon-return densities from 0 to 40,000 spawners per km² (Fig. 1). Surface sediments (0–2 cm in thickness), representing the past 5.3 ± 3.5 years of accumulation on the basis of our calculated sedimentation rates, were extracted for PCB analysis. We also measured PCB concentrations in muscle tissue of sockeye salmon ($n = 5$) to identify the signature of the source (Fig. 2a).

Surface sediments of Frazer Lake (which has annual salmon returns of 11,700 fish per km²) show patterns and concentrations of the most abundant PCBs that are similar to those carried by spawning sockeye returning to that lake (Fig. 2b). In particular, the PCB congeners designated as 101, 118, 153 and 138 + 163 were among the most abundant in both sockeye salmon and Frazer Lake sediments. More telling are the concentrations and patterns in Spiridon Lake, which receives no salmon spawners (Fig. 2c). Concentrations of PCBs in sediments from this lake are tenfold lower, and include a greater proportion of lighter congeners, which are effectively transported by air⁹.

The PCB concentration and accumulation rate in sediment correlate strongly with the density of salmon returning ($r^2 > 0.9$; Fig. 2d, e). For example, lakes with annual salmon returns of fewer than 5,000 spawners per km² contained accumulated PCB concentrations of less than 2 ng g⁻¹ (or less than 500 ng m⁻² yr⁻¹) in their surface sediments, compared with about 20 ng g⁻¹ (or about 3,700 ng m⁻² yr⁻¹) for the lake with the

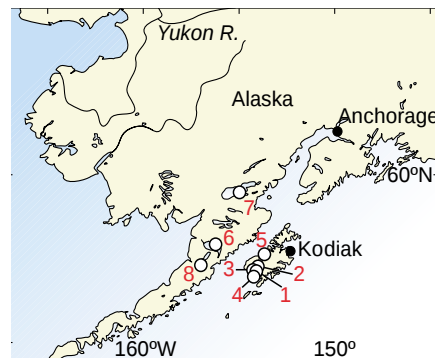


Figure 1 Location of the eight lakes where surface sediments and sockeye salmon were collected: 1, Frazer; 2, Karluk; 3, Red; 4, Olga; 5, Spiridon; 6, Becharof; 7, Iliamna; 8, Ugashik.

highest density of returning salmon (40,000 spawners per km²).

Lakes with the highest densities of spawning salmon would contribute 0.12 kg fish biomass per m², or 6,000 ng PCB m⁻² yr⁻¹. Some PCB might be lost by degradation and recycling to the water column. These values may be compared with a background PCB accumulation of about 1,000 ng m⁻² yr⁻¹ for 11 lakes distributed from mid-latitudes up to the Arctic in Canada⁹. The transport of PCBs by 40,000 salmon per km² could therefore result in a roughly sixfold increase above atmospheric loading in a remote setting, a prediction that compares well with our observations (Fig. 2d, e).

The amount of PCBs transported by sockeye salmon to these lakes is greater than the traditional assignment from atmospheric pathways. Returning sockeye salmon act as 'biological pumps' by transporting contaminants upstream, where pollutants may affect their offspring and/or predators such as bears, eagles and humans. Whether these contaminants affect juvenile salmon survival is as yet unknown, but they are suspected of causing immunosuppression¹⁰. Ironically, the marine-nutrient pump, which historically has increased successful recruitment, may now pose a risk to some of these populations.

E. M. Krümmel*, R. W. Macdonald†, L. E. Kimpe*, I. Gregory-Eaves*, M. J. Demers*, J. P. Smolt‡, B. Finney§, J. M. Blais*

*Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada
e-mail: jblais@science.uottawa.ca

†Institute of Ocean Sciences, Department of Fisheries and Oceans, Sidney, British Columbia V8L 4B2, Canada

‡Paleoecological Environmental Assessment and Research Laboratory, Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

§Institute of Marine Science, University of Alaska

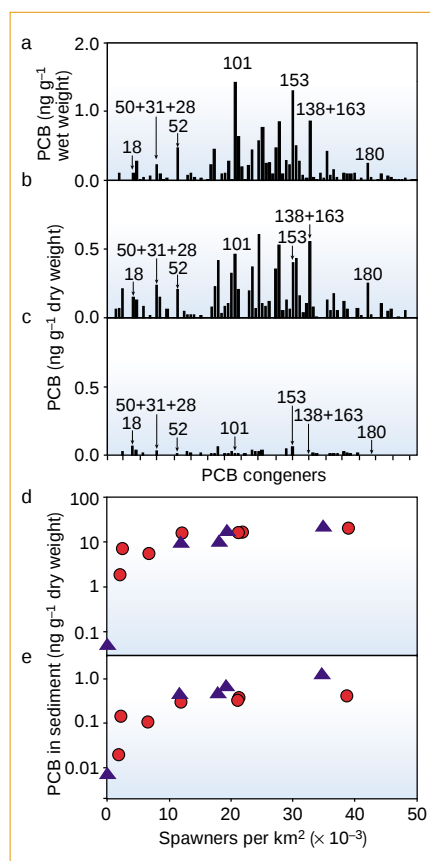


Figure 2 Surface sediments in Alaskan lakes show a similar pattern of polychlorinated biphenyl (PCB) congeners to that found in salmon returning to spawn, and sedimentary PCB concentrations are strongly correlated with the density of salmon returning. **a–c**, PCB congener patterns (numbers represent different congeners) in **a**, sockeye salmon from Frazer Lake; **b**, surface sediments from Frazer Lake (total escapement, about 11,700 spawners per km²); and **c**, surface sediments from Spiridon Lake (no anadromous salmon return). The standard deviation for salmon densities over the past ten years was on average about 50% of the mean for each lake. **d**, **e**, PCB concentrations in lake-surface sediment are shown as a function of salmon escapement density as the sum of 127 PCB congeners (**d**) and as the single PCB congener 101 (**e**). Resampling in 2002 (triangles) confirmed the 1995–1998 data (circles). Further methodological details are available from the authors.

Fairbanks, Fairbanks, Alaska 99775, USA

- Macdonald, R.W. *et al. Sci. Total Environ.* **254**, 93–234 (2000).
- Finney, B. P., Gregory-Eaves, L., Sweetman, J., Douglas, M. S. & Smol, J. P. *Science* **290**, 795–799 (2000).
- Naiman, R. J. *et al. Ecosystems* **5**, 399–417 (2002).
- Burgner, R. L. in *Pacific Salmon Life Histories* (eds Groot, C. & Margolis, L.) 1–117 (UBC, Vancouver, 1991).
- Iwata, H. *et al. Environ. Sci. Technol.* **27**, 1080–1098 (1993).
- Indian and Northern Affairs Canadian Arctic Contaminant Assessment Report: Sources, Occurrence, Trends and Pathways in the Physical Environment (Can. Ministry Pub. Works Govt Serv., Ottawa, 2003).
- Ewald, G. *et al. Arctic* **51**, 40–47 (1998).
- Blais, J. M. *et al. Environ. Toxicol. Chem.* **22**, 126–133 (2003).
- Muir, D. C. G. *et al. Environ. Sci. Technol.* **30**, 3609–3617 (1996).
- Arkoosh, M. R. *et al. J. Aquat. Anim. Health* **10**, 182–190 (1998).

Competing financial interests: declared none.

Kinematics

Wide shear zones in granular bulk flow

Granular matter does not flow homogeneously like a fluid when submitted to external stress, but usually forms rigid regions that are separated by narrow shear bands where the material yields and flows^{1–13} (examples include geological faults^{9–11}, avalanches¹² and silo discharges^{2,13}). Shear bands are narrow (five to ten grains in diameter^{1–13}) and dependent on the particle shape⁵, and often localize near a boundary^{4,8,12,13}; they hinder mixing and make grain flows difficult to predict or describe^{1–3}. Here we show that the shear zones created in the bulk of the material are wider than those

near the walls, and that their bulk velocity profiles lie on a universal curve. This finding challenges the accepted picture of shear banding in granular media.

To create tunable shear zones away from lateral boundaries, we modified a Couette cell by splitting its bottom at radius R_s and attaching the two resulting concentric rings to the inner and outer cylinder (Fig. 1a, b). The cell was filled with grains up to a height h , the outer cylinder and its co-moving ring were rotated, and the resulting flow was monitored from above by a fast CCD (charge-coupled-device) camera. We investigated the behaviour of many different types of grain, but only show results for spherical glass beads (diameter, 0.3 ± 0.1 mm).

The flow rapidly relaxed to a steady state, was purely azimuthal and was proportional to the driving rate, Ω (refs 4–8). We fixed Ω at 0.16 rad s^{-1} and measured $\omega(r)$, the dimensionless ratio of the average angular velocity, and Ω as a function of the radial coordinate. For shallow layers, a narrow shear zone developed above the split at R_s . When h was increased, this shear zone shifted away from R_s and broadened continuously and without any apparent boundary (Fig. 1b, c) — the widest zones exceeded 50 grain-diameters. The shear zone reached the inner cylinder and eventually localized there^{5,7} when the height was sufficiently large. There was, however, a substantial range of layer heights where there were wide, symmetric bulk shear zones.

After appropriate rescaling, all of these

bulk velocities collapse onto a universal curve, which is extremely well described by an error function ('erf'; Fig. 1d)

$$\omega(r) = 1/2 + 1/2 \text{erf}((r - R_s)/W) \quad (1)$$

The strain rate is therefore gaussian, and the shear zones are completely determined by their centres, R_c , and widths, W . The fit to equation (1) is just as good for particles of different size and shape. Unlike shear bands localized at walls⁵, bulk shear zones are universal — that is, they are not qualitatively influenced by the granular 'microstructure'. Removal of the inner cylinder while retaining the stationary bottom disc (dark green in Fig. 1a, b) does not affect the bulk profiles.

The evolution of the velocity profile from a step function at the bottom to an error function at the surface is reminiscent of a diffusive process along the vertical axis. However, W grows faster than $\sqrt{h}$, as diffusion would suggest, but slower than h . The shear zone's width is independent of R_s , but varies with particle size and type, hinting at a non-trivial internal length scale. By contrast, the location of the shear zone's centre, R_c , is particle-independent. Therefore, the only relevant length scales for R_c are h and R_s , and we find that the dimensionless displacement of the shear zone is well fitted by $(R_s - R_c)/R_s = (h/R_s)^{5/2}$.

Our results suggest that, for large R_s and h , the shear zones become arbitrarily broad. This raises the question of whether shear banding is intrinsic to granular matter or only occurs for particular flow geometries. Continuum theories, which should be able to describe granular shear zones, are severely constrained by the universality of the velocity profiles and the shear-zone positions, and should also incorporate the strong influence of the boundary. Our simple experimental protocol can be used to investigate unexpected regimes of granular flow, not least with a view to answering the basic question of how sand flows.

Denis Fenistein, Martin van Hecke

Kamerlingh Onnes Laboratory, Leiden University, 2300 RA Leiden, The Netherlands

e-mail: fenistein@phys.leidenuniv.nl

- Duran, J. *Sand, Powders and Grains* (Eyrolles, Paris, 1997).
- Nedderman, R. *Statics and Kinematics of Granular Materials* (Cambridge Univ. Press, Cambridge, 1992).
- Jaeger, H. M., Nagel, S. R. & Behringer, R. P. *Rev. Mod. Phys.* **68**, 1259–1272 (1996).
- Hartley, R. R. & Behringer, R. P. *Nature* **421**, 928–930 (2003).
- Mueth, D. M. *et al. Nature* **406**, 385–389 (2000).
- Howell, D. W., Behringer, R. P. & Veje, C. T. *Phys. Rev. Lett.* **82**, 5241–5244 (1999).
- Losert, W., Bocquet, L., Lubensky, T. C. & Gollub, J. P. *Phys. Rev. Lett.* **85**, 1428–1431 (2000).
- Lätzel, M., Luding, S., Herrmann, H. J., Howell, D. W. & Behringer, R. P. Preprint <http://arxiv.org/abs/cond-mat/0211274>
- Scott, D. R. *Nature* **381**, 592–595 (1996).
- Oda, M. & Kazama, H. *Géotechnique* **48**, 465–481 (1998).
- Mühlhaus, H. B. & Vardoulakis, I. *Géotechnique* **37**, 271–283 (1987).
- Daerr, A. & Douady, S. *Nature* **399**, 241–243 (1999).
- Pouliquen, O. & Gutfraind, R. *Phys. Rev. E* **53**, 552–560 (1996).

Competing financial interests: declared none.

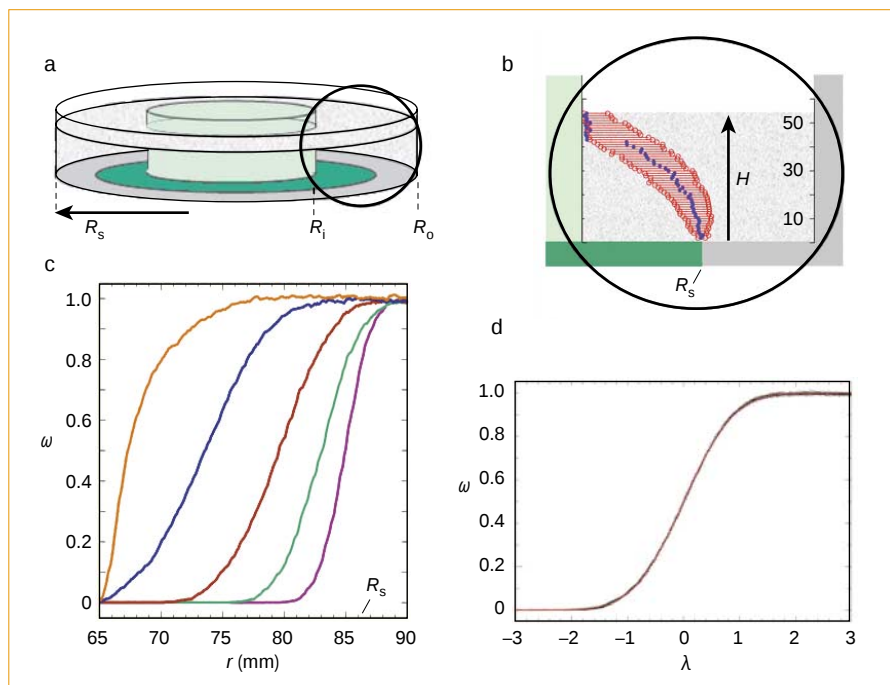


Figure 1 Creation of very wide shear zones in a flowing granular medium. **a**, Sketch of our split-bottomed Couette cell; stationary parts are shown in green. $R_i = 65$ mm, $R_o = 105$ mm; R_s can be varied. **b**, Extension of the shear zone (circled in **a**), where the angular velocity, $\omega(r)$, ranges from 0.1 to 0.9 as a function of the level-height of the grains in the cell, h , for $R_s = 85$ mm. **c**, Velocity profiles showing the evolution of the ratio of the average angular velocity, $\omega(r)$, with h ; values of h from left (in mm) are: 50, 40, 30, 20, 10. **d**, Universal bulk angular-velocity profiles obtained for different heights and plotted as a function of the rescaled radial coordinate $\lambda = (r - R_s)/W$. Red curve is an error function.

Control of leaf morphogenesis by microRNAs

Javier F. Palatnik^{1,2}, Edwards Allen³, Xuelin Wu^{2*}, Carla Schommer^{1*}, Rebecca Schwab^{1*}, James C. Carrington³ & Detlef Weigel^{1,2}

¹Department of Molecular Biology, Max Planck Institute for Developmental Biology, D-72076 Tübingen, Germany

²Plant Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

³Center for Gene Research and Biotechnology and Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331, USA

* These authors contributed equally to this work

Plants with altered microRNA metabolism have pleiotropic developmental defects, but direct evidence for microRNAs regulating specific aspects of plant morphogenesis has been lacking. In a genetic screen, we identified the *JAW* locus, which produces a microRNA that can guide messenger RNA cleavage of several *TCP* genes controlling leaf development. MicroRNA-guided cleavage of *TCP4* mRNA is necessary to prevent aberrant activity of the *TCP4* gene expressed from its native promoter. In addition, overexpression of wild-type and microRNA-resistant *TCP* variants demonstrates that mRNA cleavage is largely sufficient to restrict *TCP* function to its normal domain of activity. *TCP* genes with microRNA target sequences are found in a wide range of species, indicating that microRNA-mediated control of leaf morphogenesis is conserved in plants with very different leaf forms.

Although much is known about how organs acquire their particular fate, we are only starting to learn how organs are sculpted, even if they are just flat sheets such as wings or leaves. An elegant study recently demonstrated that making a flat organ is not a trivial problem: snapdragon leaves are normally flat, but they become crinkly in plants lacking the *CINCINNATA* (*CIN*) gene¹. In *cin* mutants, differential regulation of cell division across the leaf is disturbed, causing negative leaf curvature. *CIN* RNA itself is expressed in a dynamic pattern, in front of and perhaps overlapping the mitotic arrest zone, suggesting a direct role of *CIN* in regulating leaf morphogenesis.

Although it is unknown how expression of *CIN*, which encodes a TCP transcription factor², is regulated, a specific RNA pattern can result from differential transcription or changes in transcript

stability. A post-transcriptional mechanism that has only recently been recognized is that of plant mRNA cleavage initiated by partially or fully complementary microRNAs (miRNAs)^{3,4}. The mechanism of cleavage is similar, or identical, to cleavage guided by short interfering RNAs (siRNAs)⁵.

The double-stranded ribonucleases Dicer in animals and DICER-LIKE1 (DCL1) in plants process miRNAs—which are usually 21–22 nucleotides long—from longer precursor RNAs with fold-back structure^{4,6,7}. Additional factors required for accumulation of miRNAs include members of the Argonaute family and HEN1 protein^{8,9}. The importance of miRNAs for plant development is supported by the abnormalities seen in several mutants or transgenic plants with general defects in miRNA accumulation or activity^{9–12}. However, although biochemical studies have demon-

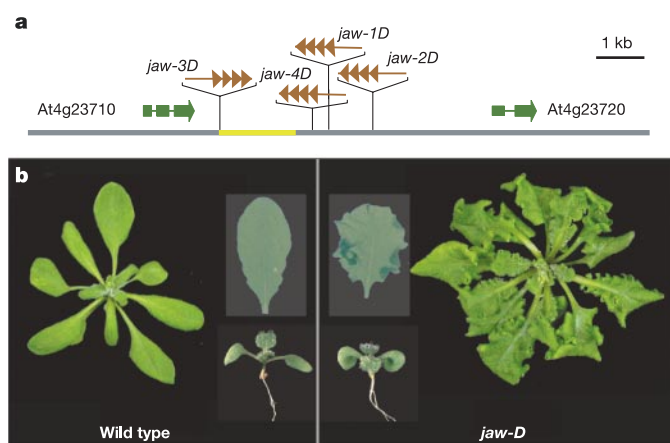
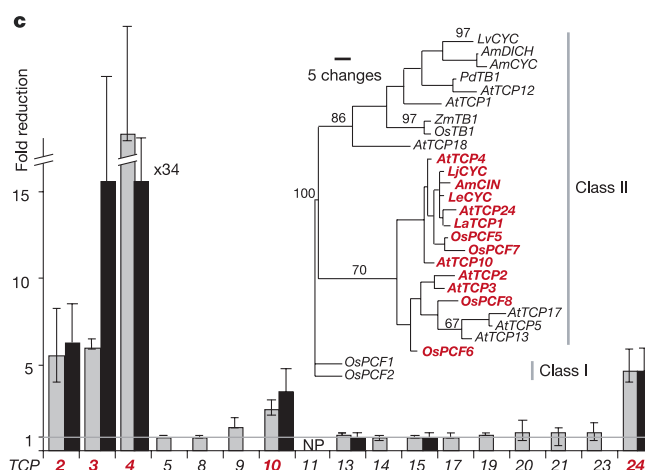


Figure 1 *jaw-D* mutations and effect of *jaw-D* on *TCP* genes. **a**, Insertion sites of *jaw-D* alleles relative to annotated genes. Triangles indicate 35S enhancers on the T-DNA vectors. The yellow bar indicates the fragment used for recapitulation (see Fig. 2b). **b**, Seedlings, individual leaves and leaf rosettes of Columbia wild-type and *jaw-1D* plants. Leaves were mounted between glass plates and illuminated from below. Dark green areas indicate overlapping leaf parts after flattening. **c**, Expression changes of *TCP* genes in *jaw-1D* estimated from Affymetrix arrays (grey bars) or from RT-qPCR (black bars). See Supplementary Information for absolute values. NP, termed 'not present' by MAS



software. Note that the Affymetrix probe sets hybridize to 3' sequences shared by full-length and cleaved transcripts, whereas the RT-qPCR products span the miRNA-guided cleavage sites and are therefore specific for the full-length transcripts. Inset shows phylogenetic tree of TCP domains, with two class I proteins^{2,38} as an out group. Red text indicates genes with miR-JAW target sequence. Am, *Antirrhinum majus* (snapdragon); At, *Arabidopsis thaliana*; La, *Lupinus albus*; Le, *Lycopersicon esculentum*; Lj, *Lotus japonicus*; Lv, *Linaria vulgaris*; Os, *Oryza sativa*; Pd, *Populus deltoides*; Zm, *Zea mays*.

strated that many plant miRNAs can guide cleavage of specific target mRNAs^{3,4,13}, the evidence for involvement of mRNA cleavage in specific biological processes has only been circumstantial.

Here, we describe an *Arabidopsis* miRNA, miR-JAW, with sequence complementarity to several *TCP* genes. MiR-JAW overexpression affects a subset of *TCP* genes, which represent the *Arabidopsis* homologues of *CIN*. The importance of miRNA-guided degradation is demonstrated with miRNA-resistant forms of target mRNAs that encode the same proteins as the wild-type genes.

Identification of the *JAW* locus and *JAW* target genes

The *JAW* locus is defined by four dominant alleles (Fig. 1a), in which viral enhancers activate an endogenous promoter that controls the transcription of an RNA lacking an annotated open reading frame¹⁴ (see Supplementary Information). We initially described *jaw-D* mutants as having serrated leaves, but a more detailed inspection revealed that the primary defect is uneven leaf shape and curvature. Unlike those of wild type, *jaw-D* leaves cannot be flattened without cutting the margins, indicating that overall gaussian curvature is not zero (Fig. 1b). All of these abnormalities are reminiscent of those seen in snapdragon *cin* mutants¹. Additional defects include coty-

ledon epinasty (Fig. 1b), a modest delay in flowering and crinkled fruits (Supplementary Information).

To understand the basis of the *jaw-D* phenotype, we analysed global expression profiles using duplicate Affymetrix arrays that represent over 24,000 annotated genes (NCBI-GEO series GSE518). Ranking by estimated reduction in expression level revealed that the top 40 genes included four *TCP* genes that are closely related to *CIN* from snapdragon. One other *TCP* gene with strong similarity to *CIN* was also downregulated (Fig. 1c), whereas the remaining *TCP* genes represented on the array showed no obvious changes. The effect of *jaw-D* on *TCP* expression was confirmed by reverse transcription followed by real-time polymerase chain reaction (Fig. 1c). The decreased expression of *CIN* homologues in *jaw-D* mutants readily explains the similar leaf phenotype of *jaw-D* and *cin* mutants.

Alignment of the *TCP* genes whose expression was reduced in *jaw-D* mutants revealed a conserved, seven-amino-acid motif outside the *TCP* domain that was missing in the *TCP* genes unaffected by *jaw-D* (Supplementary Information). The RNA sequence encoding this motif is almost invariant and has few synonymous substitutions. This RNA motif is not restricted to *CIN* homologues, and a

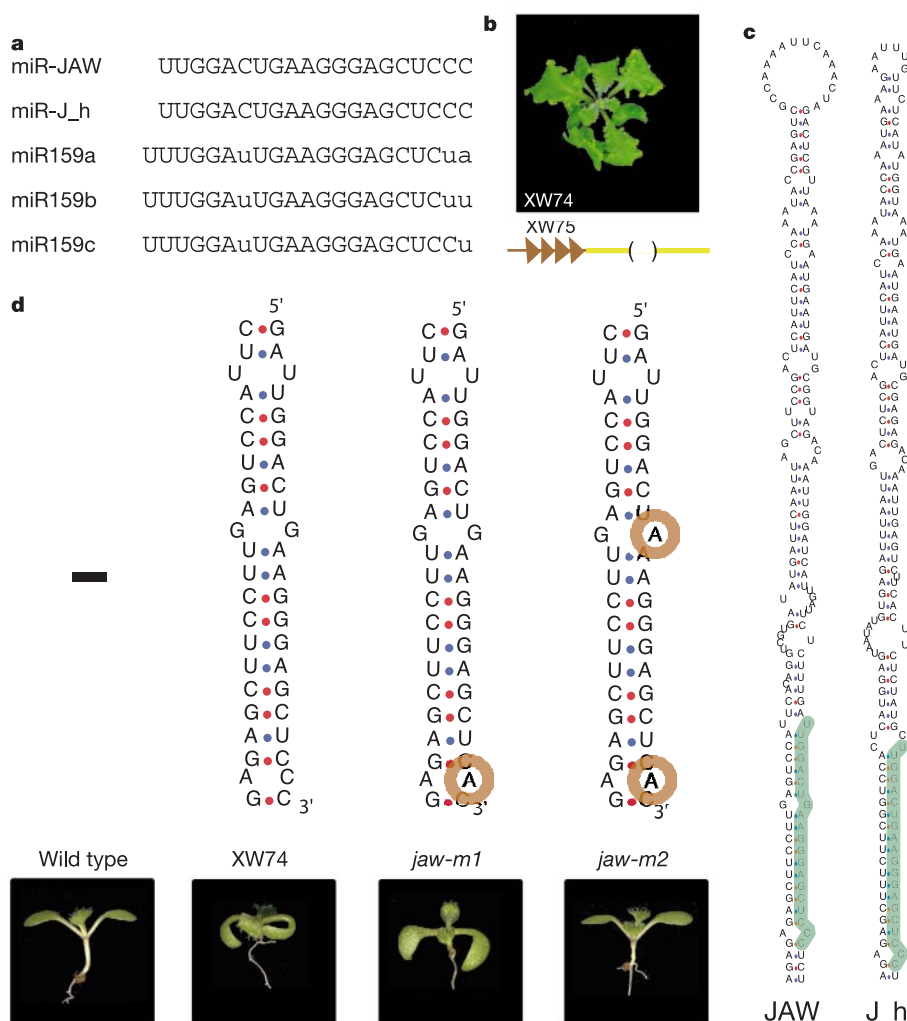


Figure 2 *JAW* encodes a microRNA. **a**, Sequence similarity of predicted miRNA from the *JAW* locus and its homologue (miR-J_h) with miR159 isoforms. **b**, The *jaw-D* phenotype was recapitulated in 106 of 151 wild-type plants transformed with pXW74, which contains a 1.4-kb fragment with adjacent 35S enhancers isolated from *jaw-3D* (see Fig. 1a). pXW75 has a 226-bp deletion and does not cause the *jaw-D* phenotype (108

transformants). **c**, Predicted fold-back structures of the miRNA precursors of *JAW* and its homologue (*J_h*) from BAC MBK23. The green lines at the right indicate the miRNAs. **d**, Effect of *JAW* point mutations in transgenic seedlings (in the context of pXW74). *jaw-m1* causes slight epinasty of cotyledons, but does not affect rosette or cauline leaves. *jaw-m2* lacks dominant *JAW* activity.

related sequence is also found in several *MYB* genes (see below). Furthermore, this motif was partially complementary to miR159 (refs 7, 9), and to a sequence at the *JAW* locus (Fig. 2a).

A miRNA encoded by *JAW*

To test the function of the motif conserved between *TCP* genes and *JAW*, we introduced a 226-base pair (bp) deletion in a genomic fragment that, when linked to 35S viral enhancers, recapitulated the *jaw-D* phenotype in transgenic plants (Figs 1a and 2b). The deleted version had lost all *JAW* activity. The deletion region can potentially give rise to an extended fold-back structure reminiscent of miRNA precursors, and a similar structure may form from a second *Arabidopsis* locus (Fig. 2c). The motif that is partially complementary to the *TCP* mRNAs, and which begins with a U like the majority of plant miRNAs^{7,9,12}, is located at the base of the putative stems. Two point mutations, in the middle and near the end of the conserved motif, were introduced into the *jaw-D* genomic fragment. Quantitative polymerase chain reaction with reverse transcription (RT-qPCR) of transgenic plants showed that production of the *JAW* RNA was not affected by these mutations (Supplementary Information). A fragment containing only the substitution at the terminal conserved position lacked most of the *JAW* activity, whereas the double mutation eliminated *JAW* activity (Fig. 2d).

We hybridized a low-molecular mass RNA blot with a 20-nucleotide probe complementary to the predicted miRNA encoded by *JAW*. This probe detected an RNA from inflorescences that co-migrated with a synthetic 20-nucleotide RNA starting with a U (Fig. 3a). It hybridized much more weakly to a larger species that co-migrated with a 21-nucleotide standard. Most miRNAs accumulate to elevated levels in plants expressing the RNA silencing suppressor P1/HC-Pro from potyviruses^{12,13,15}. This was also the case for the 20- and 21-nucleotide RNA detected by the *JAW* probe. Conversely, as with other miRNAs^{7,9,12}, these RNAs were decreased in a *dcl1* mutant. In *jaw-D*, predominantly the level of the 20-nucleotide species was increased. A probe for the related miRNA miR159 detected preferentially a 21-nucleotide species with little change in *jaw-D* (Fig. 3a). Together, these data are consistent with *JAW* encoding a 20-nucleotide miRNA, miR-JAW, which is distinct from the 21-nucleotide miR159. The 21-nucleotide species detected by the *JAW* probe may reflect the presence of a minor, modified form of miR-JAW, or cross-hybridization to miR159.

Using RT-qPCR, we found that the *MIR-JAW* precursor was more abundant in the shoot apex, inflorescence and siliques, and elevated in *jaw-D* mutants (Fig. 3b), mimicking the differential accumulation of the processed miRNA in different tissues and in *jaw-D*.

As plant miRNAs have been shown to guide target RNA cleavage^{3,4,12}, we determined whether *TCP* mRNAs were at least partially cleaved in wild-type plants. Cleavage products of all five *TCP* genes were detected using a previously described rapid amplification of cDNA ends (RACE) procedure¹² (Fig. 4a). The predominant cleavage site in each mRNA was opposite position 10 from the 5' end of miR-JAW (Fig. 4a)—which is a hallmark of miRNA- and siRNA-guided cleavage^{12,16}—thus confirming the assignment of the miR-JAW sequence. Although their low levels made detection by RNA blot difficult, fragmentation products of *TCP2* and *TCP4* appeared to be increased in *jaw-D* mutants relative to wild type (Fig. 4b). The probe sets on the Affymetrix arrays detect both full-length transcripts and 3' cleavage fragments, whereas the RT-qPCR products correspond only to uncleaved transcripts (Fig. 1a). This may explain differences in the magnitude of expression changes estimated from Affymetrix and RT-qPCR experiments.

As described above, miR-JAW also shows partial sequence complementarity to several *MYB* genes, which were predicted to be miR159 targets^{9,17}. Specific cleavage sites were detected in *MYB33* and *MYB65* mRNAs (Fig. 4a), but not in *MYB104* (not shown). The

cleavage sites are opposite position 10 from the 5' end of miR159, but only nine nucleotides downstream from the 5' end of miR-JAW. Consistent with the *MYB* genes being primarily targeted by miR159 or other, not yet identified miRNAs, the three genes were unaffected in *jaw-D* plants as determined by microarray analysis (NCBI-GEO series GSE518).

In vivo role of miRNA-guided degradation

If cleavage of *TCP* mRNAs guided by miRNAs has a role *in vivo*, we expect *TCP* RNAs to be expressed in specific patterns. RT-qPCR showed that levels of *TCP2* and *TCP4* RNA vary between different tissues. We chose to analyse *TCP4* in more detail by *in situ* hybridization because its expression was strongly affected in *jaw-D* mutants. The dynamic pattern of *TCP4* expression during embryogenesis and leaf development (Fig. 5) is consistent with *TCP4* having an instructive role in regulating the growth of cotyledons and leaves, both of which are affected in *jaw-D* mutants.

To determine the importance of miRNA-guided cleavage of *TCP4* mRNA, we engineered miRNA-resistant versions of *TCP4*. We first tested the effects of synonymous substitutions that changed the miR-JAW target sequence, but not the encoded amino acid sequence, in reconstruction experiments³ (Fig. 4a, c). In these experiments, two types of *Agrobacterium* strains were co-injected

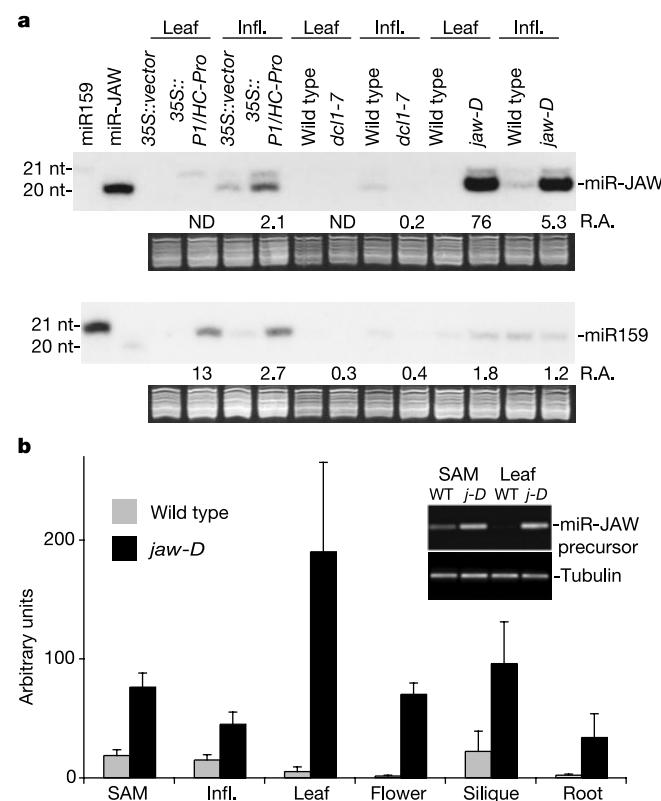


Figure 3 miR-JAW and miR159 expression. **a**, Blots of low-molecular-mass RNA extracted from leaves or inflorescences (Inf.) of different genetic backgrounds, next to synthetic miR159 (21 nucleotides) and miR-JAW (20 nucleotides). Note that there is minor cross-hybridization of the miR-JAW probe (top) and the miR159 probe (bottom). 35S::vector, empty vector control; 35S::P1/HC-Pro, 35S promoter driving expression of P1/HC-Pro. *dcl1-7* and *jaw-1D* are in different wild-type backgrounds, and different controls were used accordingly. Relative accumulation (R.A.) of miRNAs compared with the respective wild-type controls is indicated. Ethidium bromide-stained gels are shown as loading control. **b**, Expression of *MIR-JAW* precursor in different tissues of wild-type and *jaw-1D* relative to tubulin, monitored by RT-qPCR. Each bar represents the average of at least three independent samples. The inset shows examples of amplification products from shoot apical meristem (SAM) and leaf tissues separated on an agarose gel.

into a heterologous host, *Nicotiana benthamiana*. The first strain contained a construct in which either *TCP4* or a mutant derivative, *mTCP4*, was under the control of the constitutive 35S promoter. The second strain contained either an empty vector or 35S::JAW. The full-length 35S::*TCP4* transcripts were destabilized by co-injection of 35S::JAW, resulting in accumulation of cleavage products. In contrast, 35S::JAW had no effect on 35S::*mTCP4* transcript stability (Fig. 4c), demonstrating that the synonymous mutations in *mTCP4* strongly decrease susceptibility to miRNA-guided cleavage.

Next, we introduced the same mutations in the genomic context of the *TCP4* gene. Most (35 of 57) primary transformants carrying the mutated version of *TCP4* were arrested at the seedling stage (Fig. 6a), with a range of phenotypes reminiscent of embryonic patterning mutants such as *monopteros* or *cup-shaped cotyledons*^{18,19}. Common defects included fusion of the cotyledons, lack of a shoot apical meristem and overall tubular shape of the seedlings, resembling the effects of auxin inhibitors on wild-type embryos²⁰. Surviving transformants had a range of defects, including bushy rosettes and abnormal inflorescences. None of 80 primary trans-

formants containing a wild-type construct showed such defects. Together, these observations indicate that miRNA-mediated control of *TCP4* expression is essential for plant development.

Control of *TCP* mRNA abundance by miRNAs

If miRNA-guided cleavage is the major control point for regulating abundance and distribution of *TCP* genes, constitutive expression of their mRNAs should have few phenotypic effects, as long as they are susceptible to degradation. To test this hypothesis, we compared the effects of overexpressing wild-type RNAs with the effects of overexpressing RNAs with altered miRNA target sequences (Fig. 4a).

35S::*TCP2* plants had no obvious defects (100 primary transformants examined). RNA blots showed that transcripts from the 35S::*TCP2* transgene accumulated to much lower levels than transcripts from the 35S::*mTCP2* transgene (Fig. 4b), which had similar synonymous mutations to *mTCP4* (Fig. 4a). The increases in *TCP2* full-length transcripts, as detected by RT-qPCR and RNA blots, were at most twofold (Fig. 4b; see also Supplementary Information). In contrast to the wild-type construct, marked phenotypic

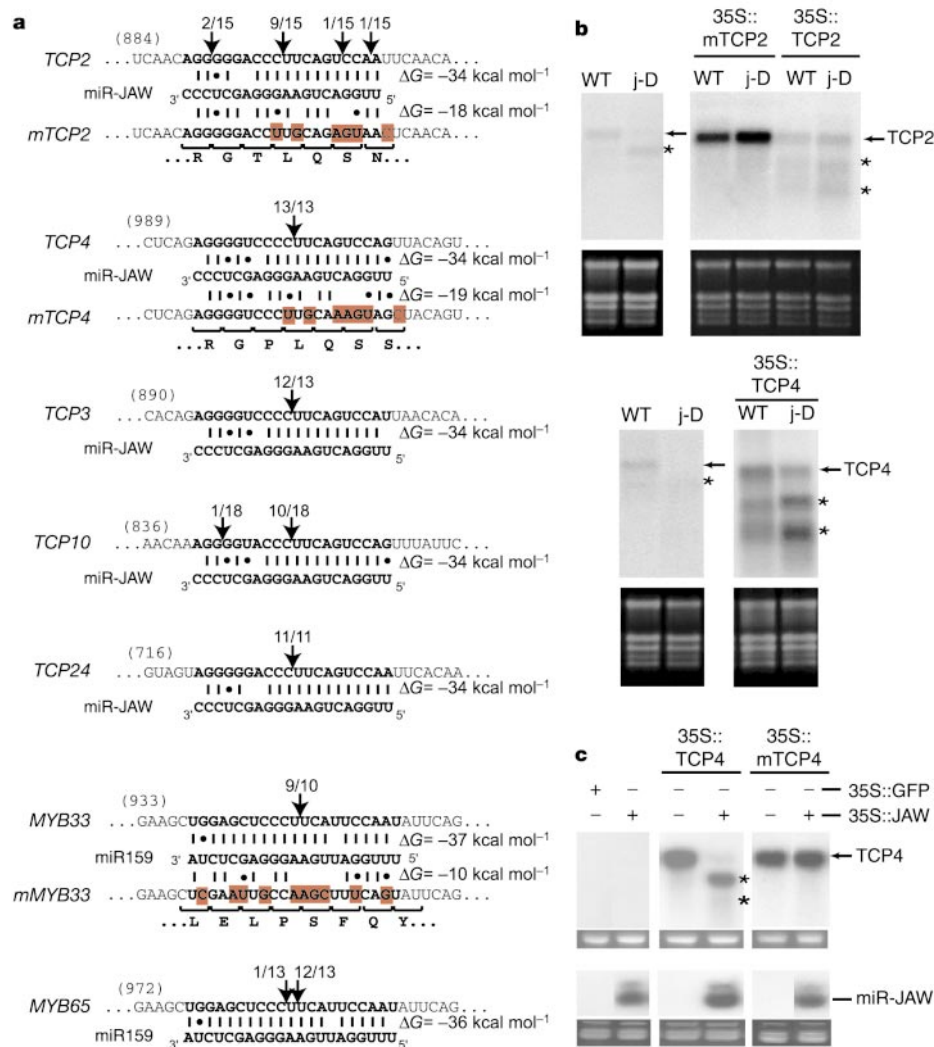


Figure 4 Identification of miRNA-guided cleavage products. **a**, Mapping of cleavage sites by 5' RACE-PCR¹². Partial mRNA sequences from five *TCP* and two *MYB* genes were aligned with miR-JAW and miR159, respectively. Free energies of duplex structures as calculated using mfold³⁹ are indicated on the right. Numbers indicate the fraction of cloned PCR products terminating at different positions. Shaded positions indicate synonymous changes introduced into miRNA target sites. **b**, Blots of high-molecular-mass RNA from different *Arabidopsis* lines. Arrows indicate intact mRNA molecules;

asterisks indicate fragments resulting from miRNA-guided cleavage. The ratio of full-length transcripts to cleavage products was always lower in *jaw-1D*. **c**, Blots of high-molecular-mass (top) and low-molecular-mass (bottom) RNA from *N. benthamiana* leaves co-infiltrated with different *Agrobacterium* strains. The 20-nucleotide RNA corresponding to miR-JAW was detected only in leaves injected with 35S::JAW (pXW74). A less prominent 21-nucleotide species was also detected. Ethidium bromide-stained gels are shown as loading controls.

effects were frequent with *35S::mTCP2*, consistent with the substantially higher levels of full-length transcripts (Fig. 4b; see also Supplementary Information). Many (30 of 71) *35S::mTCP2* plants had longer hypocotyls than wild type, they were smaller and greener, and their apical dominance was reduced (Fig. 6b).

A similar picture was seen with *TCP4*. Most (63 of 76) primary *35S::TCP4* transformants were phenotypically normal, with the remainder having slightly longer hypocotyls and smaller leaves than wild type (not shown). In contrast, most (55 of 68 of primary transformants) of the *35S::mTCP4* plants had marked pattern defects (Fig. 6c), similar to plants with the mutant *TCP4* form under control of the endogenous promoter, but at a slightly higher frequency (81% compared with 61% of transformants).

From experiments with mutated *MYB33* we conclude that miRNA regulation of *MYB33* is important as well. Whereas *35S::MYB33* had no obvious effect on development, *35S::mMYB33* caused leaves to curl upwards in many (39 of 63) primary transformants (Fig. 6d).

TCP mRNA degradation causes *jaw-D* phenotypes

To confirm that degradation of target mRNAs was indeed the cause for *jaw-D* phenotypes, we introduced constructs expressing wild-type and miRNA-resistant forms of target genes into *jaw-D* mutants. Constitutive expression of wild-type *TCP2* or *TCP4* partially rescued the leaf defects of *jaw-D* mutants. Leaves were less warped and elliptical than those of the parental line, but were still distinct from those of wild type (Fig. 6e). The suppression of the *jaw-D* phenotype was paralleled by an increase in *TCP2* or *TCP4* full-length RNA (Fig. 4b). RT-qPCR confirmed that rescue was not due to co-suppression and thereby reduced expression of the *MIR-JAW* precursor (Supplementary Information).

The mutant version of *TCP2* was able to rescue the leaf shape and curvature defects of *jaw-D* more extensively than *35S::TCP2* (Fig. 6d). The level of full-length *mTCP2* mRNA from the transgene was similar in wild-type and *jaw-D* backgrounds, and substantially higher than either the endogenous mRNA or the *TCP2* mRNA produced from the non-mutated *35S::TCP2* construct (Fig. 4b). The correlation between rescue and amount of full-length transcripts was further confirmed by RT-qPCR (Supplementary Information). The *35S::mTCP4* construct caused the same severe defects in *jaw-D* as in a wild-type background (not shown). Because these plants did not produce clearly identifiable cotyledons or leaves, rescue of the *jaw-D* phenotype could not be assessed. Expression of *mTCP4* from

the *TCP4* promoter had less extreme effects than *35S::mTCP4* (see above), and the *jaw-D* leaf phenotype was suppressed in 11 of 15 surviving primary transformants. Notably, these plants showed a new phenotype: rounder leaves (Supplementary Information).

Constitutive expression of mutated *MYB33* in *jaw-D* mutants caused an additive phenotype (not shown), consistent with the observation that *MYB33* RNA was not reduced in *jaw-D* mutants, and further supporting the suggestion that *MYB33* is primarily targeted by an miRNA other than miR-JAW, probably miR159.

MIR-JAW and miRNA target-site selection

Expressed sequence tags (ESTs) with similarity to the *MIR-JAW* precursor were found in several species, including wheat and soybean (Supplementary Information). All ESTs are much longer than the predicted fold-back structure in *MIR-JAW*, similar to the approximately 500-base *JAW* transcript from *Arabidopsis*¹⁴, suggesting that sequences flanking the fold-back structure are important for miRNA processing, even though there is little similarity at the primary sequence level.

Database searches also identified homologues of candidate miR-JAW targets among *TCP* genes in several species. All 11 proteins encoded by genes containing an miRNA target site belong to a monophyletic clade within the class II TCP subgroup (Fig. 1c), indicating that the miRNA target sequence arose only once among the *TCP* genes.

When we compared divergence of coding sequences for TCP domains and miRNA target motifs, we found similarly low rates of non-synonymous changes ($K_A = 0.17$ and 0.12 , respectively), but a much higher rate of synonymous changes ($K_S = 1.22$) in the TCP domain than in the miRNA target motif ($K_S = 0.35$), indicating that the selective forces acting on the miRNA target motif are different from those that shape the TCP domain (see Supplementary Information for details).

Discussion

MicroRNAs and siRNAs, which have been found in all complex eukaryotes examined, regulate the activity of target genes through at least two distinct mechanisms. In both plants and animals, siRNAs with perfect complementarity trigger degradation of their target mRNAs^{4,21,22}. In contrast, the predominant mode of miRNA action in animals seems to be translational control, as originally shown for the prototypical miRNAs produced by the *let-7* and *lin-4* genes of *Caenorhabditis elegans*^{23–25}, and more recently for *bantam* in *Dro-*

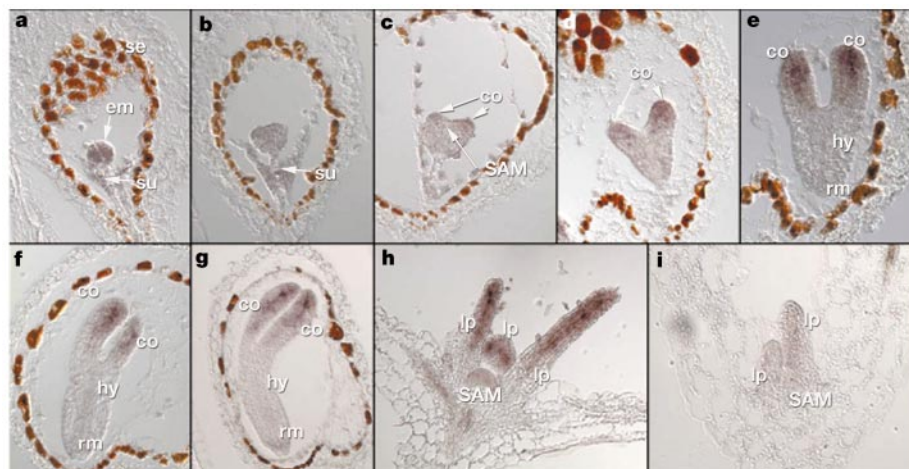


Figure 5 Expression pattern of *TCP4* mRNA. **a–g**, *In situ* localization of *TCP4* mRNA in wild-type embryos of progressive age. From an initially uniform pattern during the globular stage (**a**), *TCP4* RNA becomes gradually restricted to emerging cotyledons during the heart stage (**d**). During the torpedo stage, *TCP4* RNA becomes localized to subepidermal

regions of the cotyledons, with highest levels near the tip (**e–g**). **h**, A similar pattern as in cotyledons is seen in young leaves of a wild-type seedling. **i**, *TCP4* RNA signal is reduced in *jaw-D* seedlings. co, cotyledon; em, embryo proper; hy, hypocotyl; lp, leaf primordium; rm, root meristem; SAM, shoot apical meristem; se, seed coat; su, suspensor.

*sophila melanogaster*²⁶ and miR-23 in human cells²⁷. Although naturally occurring animal miRNAs are able to target artificial, perfectly complementary mRNAs for destruction²⁸, their normal targets have only limited complementarity, which apparently prevents degradation of the target RNAs.

In plants, it is unknown whether miRNAs are involved in translational control, but there are several examples of mRNAs that are cleaved in positions complementary to endogenous miRNAs^{3,4,12,13}. Circumstantial evidence for this mechanism being important in regulating accumulation of specific mRNAs came initially from an analysis of dominant mutations at the *PHV* locus. These mutations affect RNA cleavage in a cell-free system⁴ and cause aberrant accumulation of *PHV* mRNA *in vivo*²⁹; however, because they also change the amino acid sequence of a potential regulatory motif predicted to be involved in sterol- or lipid-binding, it was unclear whether the primary defect is indeed in aberrant mRNA accumulation, and not, as originally proposed²⁹, due to a change in protein function.

We identified a new miRNA, miR-JAW, using a genetic screen, along with a new set of miRNA targets related to the *CIN* gene¹ of snapdragon. These targets constitute a subset of *TCP* genes, which, in contrast to several other transcription factor genes with potential roles in development, had not been previously recognized as candidate miRNA targets. We have shown that miRNA-directed cleavage of *TCP* mRNAs is not only required for normal develop-

ment, but that it is also the main control point for regulating mRNA accumulation.

CIN mRNA is normally downregulated in non-dividing, differentiating cells. Developing leaves of *cin* mutants therefore suffer from a delay in cell division arrest (and subsequent differentiation), leading to accumulation of excess cells at the periphery and a crinkled leaf phenotype¹. Thus, miRNA-mediated control of *CIN* and other target *TCP* genes is probably required for proper timing of the transition between cell division and differentiation in developing leaves. Both the *MIR-JAW* precursor and the miRNA target motif in *TCP* genes are found across a wide range of flowering plants, even though these have very different leaf morphologies and modes of leaf development. This observation supports a conserved role of the *CIN*-class of *TCP* genes in controlling leaf morphogenesis.

Although the growth-arrest phenotype of dominant *TCP4* mutants is consistent with a role of *Arabidopsis* *TCP* genes in differentiation, some of the dominant *TCP2* phenotypes, such as elongated hypocotyls or reduced apical dominance, are not readily explained by growth inhibition. Conversely, *jaw-D* mutants have phenotypes in addition to crinkly leaves, such as late flowering, suggesting additional roles for *TCP* genes and miR-JAW targets. A further layer of complexity is added by the fact that the *Arabidopsis* genome contains at least two regions with the potential to produce miR-JAW, as well as several potential precursors for a family of three almost identical miRNAs (miR159a–c) that are related to miR-JAW. In contrast to miR-JAW, miR159 is predicted to preferentially target three *MYB* genes, as miR159 shows better complementarity with the *MYB* than the *TCP* genes.

Our finding that only overexpression of an miRNA-resistant form of *MYB33* has major phenotypic consequences provides strong evidence for miRNA regulation of *MYB33*. It does not discriminate between miRNA-controlled RNA cleavage and translational repression, which leaves several possible explanations for the absence of an effect of *jaw-D* on *MYB* RNA expression. Our favourite model is that *MYB* genes are not cleaved by miR-JAW, but by other miRNAs such as miR159. Alternatively, translational regulation of *MYB* genes by miRNAs may be more important than mRNA cleavage. These observations highlight the importance of examining predicted targeting events *in vivo*.

The complex relationships between related miRNAs on the one hand and overlapping sets of potential targets on the other hand indicates that the miRNA regulatory network is similarly intricate compared with transcriptional networks in which families of related DNA-binding proteins control overlapping sets of target genes. Global transcript analysis, together with miRNA overexpression and the engineering of miRNA-resistant target genes, as implemented in our study, should be powerful tools for dissecting the miRNA regulatory network in plants.

Note added in proof: Overexpression experiments suggest that some plant miRNAs act also through translational control⁴⁰. □



Figure 6 Effects of miRNA-resistant transgenes and trans-complementation of *jaw-1D*. **a**, Different patterning defects in seedlings transformed with a genomic copy of *TCP4* containing point mutations in the miRNA target site (see Fig. 4a). **b**, Plants that overexpress mutated *TCP2* (*35S::mTCP2*); note the elongated hypocotyls shown in the inset (wild type to the left). **c**, Patterning defects in *35S::mTCP4* seedlings; compare to **a**. **d**, Effects of mutated *MYB33* (*35S::mMYB33*) on rosette morphology. **e**, Partial to complete complementation of the *jaw-D* leaf phenotype by overexpressing wild-type *TCP2* and *TCP4*, or mutated *TCP2*.

Methods

Plant material

Plants were grown in long days (16 h light/8 h dark) under fluorescent lights at 23 °C. *jaw-1D*, *jaw-2D*, *dcl1-7* and *35S::P1/HC-Pro* plants have been described^{10,12,14}. *jaw-3D* and *jaw-4D* were isolated using pSKI015 (ref. 14). Transgenic plants were generated as described³⁰.

Transgenes

The genomic fragment used in pXW74 was obtained by plasmid rescue¹⁴. *TCP* and *MYB* complementary DNAs were amplified from a Columbia cDNA library (gift of P. Wiggle). Mutated versions of *TCP* genes and *MYB33* were generated by PCR. Promoter sequences of *TCP4* were isolated from genomic DNA by PCR. All amplification products were verified by sequencing. Primer sequences are available on request.

Microarray analysis

RNA was extracted using the Plant RNeasy Mini kit (Qiagen). Double-stranded cDNA was synthesized from 10 µg total RNA using the Superscript Choice System (Invitrogen) and

an oligonucleotide dT-T7 primer (Genset). cDNA was converted to biotinylated cRNA using the BioArray High Yield Transcript Labelling kit (Enzo). cRNA was cleaned with RNeasy columns (Qiagen) according to the manufacturer's protocol, with the following modifications. First, the cRNA was passed through the column twice to increase binding. Second, the eluate was re-applied to the column once to increase yield. Usually, 50–100 µg cRNA were obtained. Twenty micrograms of concentration-adjusted cRNA were fragmented and hybridized to GeneChip arrays according to the manufacturer's protocol (Affymetrix). For washing and staining, protocol EukGE-WSv4 was used. Expression values were estimated using Microarray Suite v5.0 (Affymetrix) or robust multi-array analysis³¹.

Real-time RT-PCR

Reverse transcription was performed with 1 µg total RNA, using a kit (Promega). PCR was carried out in the presence of the double-strand DNA-specific dye SYBR green (Molecular Probes). Amplification was monitored in real time with the Opticon Continuous Fluorescence Detection System (MJR). Dual-labelled fluorogenic probes were used for TaqMan assays³². Primer sequences are available on request.

Small RNA isolation and blot analysis

Total RNA was isolated from fully expanded leaves and inflorescence clusters (containing stage 1–12 flowers; ref. 33) using Trizol reagent (Invitrogen). Low-molecular-mass RNA was separated from high-molecular-mass RNA using Qiagen Midi preps, and resolved by 17% PAGE under denaturing conditions³. A total of 250 pg synthetic RNA oligonucleotides that were 5' phosphorylated (Dharmacon Research) were included as size standards. Blots were hybridized using end-labelled oligonucleotide probes³. Hybridization signal was measured using an InstantImager (PerkinElmer).

Cleavage site mapping of miRNA targets

An RNA ligase-mediated 5' RACE procedure was used to map the cleavage site of miRNA targets as described¹². Total RNA from 4-week-old *Arabidopsis* (Col-0) leaves and inflorescences was used for poly(A)⁺ mRNA preparation. After amplification, 5' RACE products were gel-purified and cloned, and at least ten independent clones were sequenced. Primer sequences are available on request.

In situ hybridization

Previously published methods³⁰ were used with modifications, including the use of an automated embedding system (Leica).

Phylogenetic analysis

Sequences for full-length *TCP* genes were identified from GenBank by BLAST analysis. The gene names and their GenBank accession numbers are provided in Supplementary Information. Protein and *MIR-JAW* precursor sequences were aligned using Clustal³⁴. *TCP* nucleotide sequences were aligned manually to match protein alignment. Phylograms were generated with PAUP* (ref. 35), using a heuristic search with Multrees option and the TBR swapping algorithm, using 100 replicates (bootstrap values shown in Fig. 1c). A total of 267 trees were generated by maximum parsimony. Bayesian analysis was performed using MrBayes³⁶, with the chains run for 2,000,000 generations. For nucleotide analysis, the third positions of codons were excluded. Synonymous and non-synonymous site changes between pairs of sequences were calculated using DnaSP³⁷. A 21-nucleotide region encompassing the miRNA target sequence and corresponding 60-nucleotide regions from genes lacking the target motif were selected for analysis.

Received 21 June; accepted 31 July 2003; doi:10.1038/nature01958.

Published online 20 August 2003.

- Nath, U., Crawford, B. C., Carpenter, R. & Coen, E. Genetic control of surface curvature. *Science* **299**, 1404–1407 (2003).
- Cubas, P., Lauter, N., Doebley, J. & Coen, E. The *TCP* domain: a motif found in proteins regulating plant growth and development. *Plant J.* **18**, 215–222 (1999).
- Llave, C., Xie, Z., Kasschau, K. D. & Carrington, J. C. Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297**, 2053–2056 (2002).
- Tang, G., Reinhart, B. J., Bartel, D. P. & Zamore, P. D. A biochemical framework for RNA silencing in plants. *Genes Dev.* **17**, 49–63 (2003).
- Plasterk, R. H. RNA silencing: the genome's immune system. *Science* **296**, 1263–1265 (2002).
- Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**, 834–838 (2001).
- Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B. & Bartel, D. P. MicroRNAs in plants. *Genes Dev.* **16**, 1616–1626 (2002).
- Carmell, M. A., Xuan, Z., Zhang, M. Q. & Hannon, G. J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
- Park, W., Li, J., Song, R., Messing, J. & Chen, X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12**, 1484–1495 (2002).
- Jacobsen, S. E., Running, M. P. & Meyerowitz, E. M. Disruption of an RNA helicase/RNase III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* **126**, 5231–5243 (1999).
- Golden, T. A. *et al.* *SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY*, a Dicer homolog,

is a maternal effect gene required for embryo development in *Arabidopsis*. *Plant Physiol.* **130**, 808–822 (2002).

- Kasschau, K. D. *et al.* P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. *Dev. Cell* **4**, 205–217 (2003).
- Xie, Z., Kasschau, K. D. & Carrington, J. C. Negative feedback regulation of *Dicer-Like1* in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr. Biol.* **13**, 784–789 (2003).
- Weigel, D. *et al.* Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**, 1003–1013 (2000).
- Mallory, A. C., Reinhart, B. J., Bartel, D., Vance, V. B. & Bowman, L. H. A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. *Proc. Natl Acad. Sci. USA* **99**, 15228–15233 (2002).
- Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877–6888 (2001).
- Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
- Mayer, U., Torres Ruiz, R. A., Berleth, T., Miséra, S. & Jürgens, G. Mutations affecting body organization in the *Arabidopsis* embryo. *Nature* **353**, 402–407 (1991).
- Aida, M., Ishida, T., Fukaki, H., Fujisawa, H. & Tasaka, M. Genes involved in organ separation in *Arabidopsis*: an analysis of the *cup-shaped cotyledon* mutant. *Plant Cell* **9**, 841–857 (1997).
- Hadfi, K., Speth, V. & Neuhaus, G. Auxin-induced developmental patterns in *Brassica juncea* embryos. *Development* **125**, 879–887 (1998).
- Hamilton, A. J. & Baulcombe, D. C. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**, 950–952 (1999).
- Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
- Lee, R. C., Feinbaum, R. L. & Ambros, V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843–854 (1993).
- Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
- Reinhart, B. J. *et al.* The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
- Brenneke, J., Hipfner, D. R., Stark, A., Russell, R. B. & Cohen, S. M. *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* **113**, 25–36 (2003).
- Kawasaki, H. & Taira, K. *Hes1* is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells. *Nature* **423**, 838–842 (2003).
- Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
- McConnell, J. R. *et al.* Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. *Nature* **411**, 709–713 (2001).
- Weigel, D. & Glazebrook, J. *Arabidopsis: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002).
- Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15, doi:10.1093/nar/gng015 (2003).
- Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).
- Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in *Arabidopsis*. *Plant Cell* **2**, 755–767 (1990).
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
- Swofford, D. L. PAUP: A computer program for phylogenetic inference using maximum parsimony. *Gen. Physiol.* **102**, 9A (1993).
- Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
- Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).
- Citerne, H. L., Luo, D., Pennington, R. T., Coen, E. & Cronk, Q. C. A phylogenomic investigation of *CYCLOIDEA-like TCP* genes in the Leguminosae. *Plant Physiol.* **131**, 1042–1053 (2003).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
- Chen, X. A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* published online 31 July 2003 (doi:10.1126/science.1088060).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank I. Puga-Gonzalez for assistance and V. Ambros, S. Balasubramanian, J. Chory, M.-C. Kim, J. Lohmann, Y. Kobayashi and J. Spatafora for advice and discussion. This work was supported by fellowships from CONICET and Human Frontier Science Program Organization to J.F.P., from Life Sciences Research Foundation/US Department of Energy to X.W.; by grants from NSF and NIH to J.C.C. and NIH to D.W.; and by the Max Planck Society. D.W. is a Director of the Max Planck Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W. (weigel@weigelworld.org) or J.C.C. (carrington@orst.edu).

The formation of cluster elliptical galaxies as revealed by extensive star formation

J. A. Stevens¹, R. J. Ivison¹, J. S. Dunlop², Ian R. Smail³, W. J. Percival², D. H. Hughes⁴, H. J. A. Röttgering⁵, W. J. M. van Breugel⁶ & M. Reuland⁶

¹Astronomy Technology Centre, Royal Observatory, and ²Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK

³Institute for Computational Cosmology, University of Durham, South Road, Durham, DH1 3LE, UK

⁴Instituto Nacional de Astrofísica, Óptica y Electrónica, Apartado Postal 51 y 216, 72000 Puebla, Mexico

⁵Leiden Observatory, PO Box 9513, 2300 Leiden, The Netherlands

⁶Institute of Geophysics and Planetary Physics, Lawrence Livermore National Laboratory, PO Box 808, Livermore, California 94459, USA

The most massive galaxies in the present-day Universe are found to lie in the centres of rich clusters. They have old, coeval stellar populations suggesting that the bulk of their stars must have formed at early epochs in spectacular starbursts¹, which should be luminous phenomena when observed at submillimetre wavelengths². The most popular model of galaxy formation predicts that these galaxies form in proto-clusters at high-density peaks in the early Universe³. Such peaks are indicated by massive high-redshift radio galaxies⁴. Here we report deep submillimetre mapping of seven high-redshift radio galaxies and their environments. These data confirm not only the presence of spatially extended regions of massive star-formation activity in the radio galaxies themselves, but also in companion objects previously undetected at any wavelength. The prevalence, orientation, and inferred masses of these submillimetre companion galaxies suggest that we are witnessing the synchronous formation of the most luminous elliptical galaxies found today at the centres of rich clusters of galaxies.

Although existing submillimetre studies of high-redshift radio galaxies (hereafter HzRGs) have shown that their star-formation rates are large enough to build a massive galaxy in <1 Gyr (refs 2, 5–7), they have provided no information on the spatial extent of this emission or on the star-formation activity in their environments. We have therefore mapped a sample of seven objects with redshifts ranging from 2.2 to 4.3 at a wavelength of $850\ \mu\text{m}$ with the Submillimetre Common-User Bolometer Array (SCUBA)⁸ on the James Clerk Maxwell Telescope (JCMT). The targets were chosen from those sources found to be submillimetre bright in the previous SCUBA surveys of HzRGs^{6,7}. Our new maps illustrate the distribution of dust-radiated emission in and around the HzRGs on scales from $5''$ to $160''$, or 30 kpc to 1 Mpc. We illustrate the seven submillimetre maps from this survey in Fig. 1; the orientation of the radio jets of each HzRG is represented by yellow tick marks on these maps.

One of the most striking aspects of the submillimetre maps is that the dust emission from the central radio galaxy is resolved in at least five of the seven sources—even with the coarse beam of the JCMT. In Fig. 2 and Table 1 we present simple two-dimensional gaussian fits to the data which, while not giving a true reflection of the physical situation, at least provide a quantitative measure of the spatial extent of the dust emission. This emission is sometimes in the form of several partially resolved or merged clumps (typified by 8C1909+722), sometimes in an apparently smoother distribution (for example, 4C60.07), and is more extended than the radio emission in most cases.

The extent of the dust emission ranges from 50 to 250 kpc, a physically interesting size because (1) the corresponding half-light

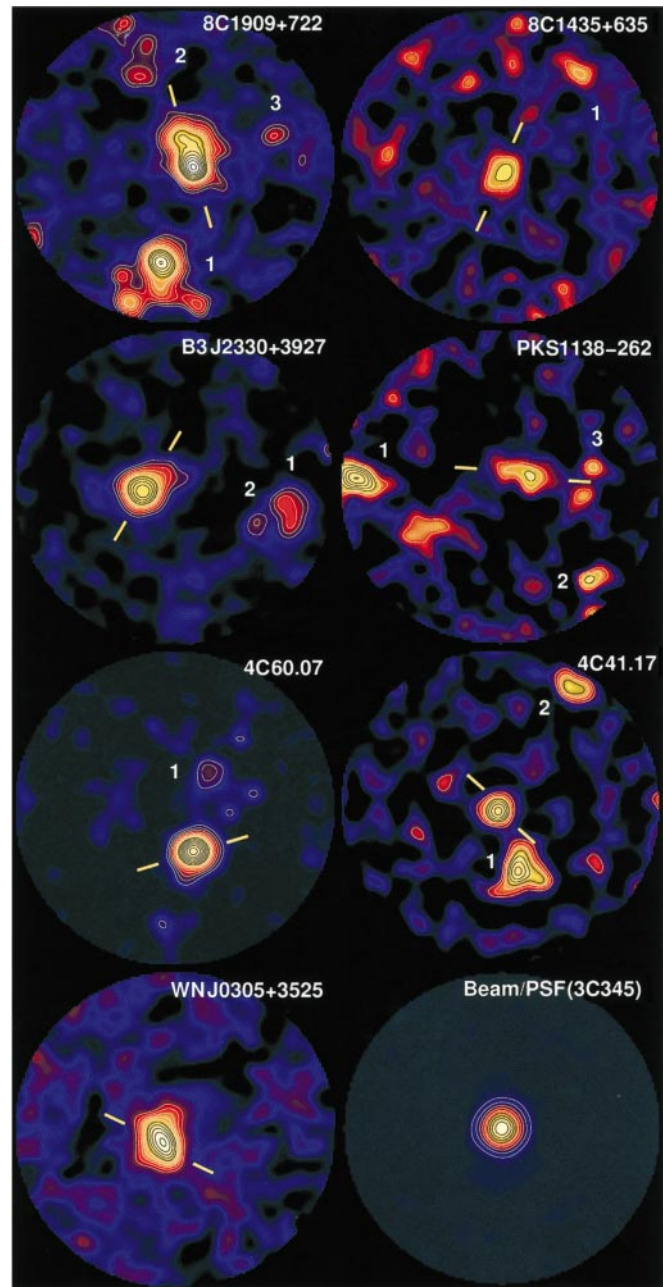


Figure 1 Continuum emission from dust in and around seven high-redshift radio galaxies. The data for 4C41.17 have been published previously²⁷. Data were obtained using the SCUBA⁸ submillimetre camera on the JCMT in its dual-wavelength mapping mode during the period 1998–2001 in the top 20 percentiles of Mauna Kea weather conditions. The beam, after smoothing to a FWHM of $14.4''$ at $850\ \mu\text{m}$, is shown bottom right in the form of a map of the blazar, 3C345, with contours at $-40, -30, -10, +10, +30, +50, +70$ and $+90\%$ of the central peak. The JCMT secondary mirror was chopped and nodded by $30''$ in RA, resulting in the $-0.5/+1.0/-0.5$ beam profile. After all overheads, 30–40 ks were spent integrating on each of the fields, a total of ~ 100 h before overheads, split down into 1,280-s integrations separated by checks on focus, pointing accuracy and atmospheric opacity. All images have been cleaned with the smoothed beam²⁷. Submillimetre contours are shown at $-3, 3, 4, 5, 6, 8$ and $10 \times \sigma$, where σ includes the contribution due to sources below the detection threshold and to the chopping and nodding procedure, and ranges from 1 to $2\ \text{mJy beam}^{-1}$. The galaxy with the most compact appearance, 4C41.17, actually subtends $\sim 7.5''$ FWHM when deconvolved from the beam. The yellow tick marks around the central radio galaxies show the direction of the kiloparsec-scale radio jets. PSF, point spread function.

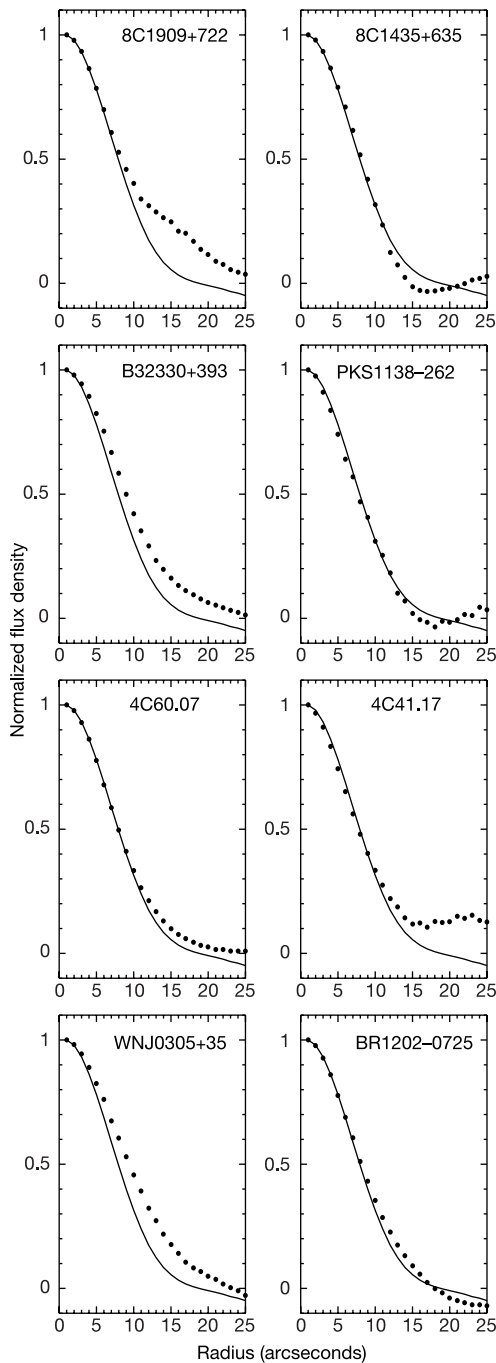


Figure 2 Submillimetre extent of the radio galaxies. Radial profiles of the dust emission compared with that of the beam (solid line). An accurate beam profile or PSF was measured with a ~ 1 -h mapping observation of the blazar 3C345 which had an $850\text{-}\mu\text{m}$ flux density of $>2\text{ Jy}$ at the time. The image (shown in Fig. 1) has a FWHM of $14.4''$. All subsequent size measurements are deconvolved from the PSF (see Table 1). The reality of the extended nature of the dust emission is established from observations of the quasar BR1202–0725, using the same technique and integration times used for the radio galaxies to mimic any systematic errors. BR1202–0725 serves as a particularly conservative comparison object—its dust continuum was resolved²⁸ by the IRAM interferometer into two components separated by $4''$ (PA $120 \pm 10^\circ$). As expected, it appears relatively compact ($6.5'' \times 5.5''$ at PA 6°) in our submillimetre maps, and its radial profile is much more similar to that of the beam than to those of the HzRGs. We conclude that our observing procedure does not produce spurious morphologies, and that the extended nature of the submillimetre emission from the radio galaxies is real, occurring on scales that are typically much greater than $4''$.

radii (30–150 kpc) are equivalent to those of the brightest cluster galaxies in the local Universe⁹, and (2) gas-dynamical simulations of major galaxy mergers predict that star formation should peak when the galaxies are separated on approximately this scale¹⁰. Higher-resolution, high signal-to-noise (S/N) millimetre/submillimetre imaging observations are required to investigate these possibilities.

A second striking feature of our SCUBA maps is that several of the fields also contain serendipitous detections of new, luminous submillimetre galaxies. In Table 1 we list the properties of those companion galaxies with peak $S/N > 4$. Several of these companions are, like the HzRGs, resolved at $850\text{ }\mu\text{m}$. The source density of companions is also higher than found in ‘blank-field’ surveys; current submillimetre number counts¹¹ indicate that we should find, on average, about one random submillimetre source per SCUBA field at a flux level of $S_{850} > 5\text{--}6\text{ mJy}$ compared to our detection of approximately twice this number (Table 1). Furthermore, the 4C41.17 and 8C1909+722 fields each contain a companion source with $S_{850} > 20\text{ mJy}$. By comparison, the biggest blank-field survey conducted until now¹², which has an area about seven times that of our own, contains no robust sources with $S_{850} > 15\text{ mJy}$.

Arguably the most surprising feature of the maps is the observed alignments between the HzRG radio axis and (1) the submillimetre

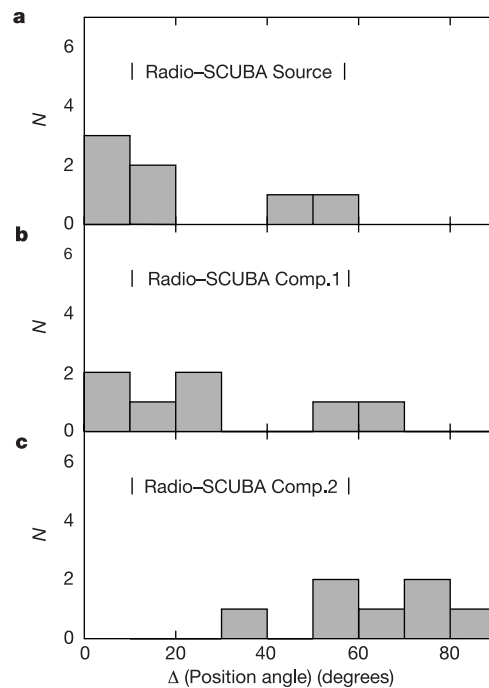


Figure 3 Observed alignment effects. **a**, **b**, The histograms show the offset in degrees between the kiloparsec scale radio structure of the HzRG and the position angle of the submillimetre source (**a**), the position angle defined by the projected vector joining the radio source to the brightest, spatially distinct submillimetre source (**b**), and the position angle defined by the projected vector joining the radio source to the second brightest, spatially distinct submillimetre source (**c**). In four cases these submillimetre sources are the companions listed in Table 1. In the remaining three cases one or both of the companions has an S/N ratio of 2 to 4; although these sources are not listed in Table 1, we use them in the analysis for completeness. We have applied the Kolmogorov Smirnov test to assess whether the apparent alignment effect seen in the first two of these plots is indeed statistically significant. Despite the small-number statistics, we find that both of these apparent alignment effects are significant at the 2σ level (as compared to a random distribution; $P = 0.04$ and $P = 0.05$ respectively), whereas the corresponding result for the third distribution is not significant ($P = 0.3$).

Table 1 **Target fields**

Target name	Coordinates (J2000)		z	S_{peak} (mJy)	S_{total} (mJy)	$M_d(M_\odot)$	FWHM at PA(°)	Radio size at PA(°)
	RA	Dec.						
8C1909+722	19 h 08 min 23.30 s	+72° 20' 10.4"	3.54	20.5 ± 1.2	34.9 ± 3.0	1.2×10^8	26 × 9 at 18°	14 at 15°
1	19 h 08 min 27.47 s	+72° 19' 28.0"	...	17.6 ± 1.2	23.0 ± 2.5	...	12 × 8 at 169°	...
2	19 h 08 min 29.31 s	+72° 20' 49.6"	...	6.5 ± 1.2	8.7 ± 2.4	...	...	...
3	19 h 08 min 16.12 s	+72° 20' 24.0"	...	6.5 ± 1.2	4.3 ± 2.1	...	...	...
8C1435+635	14 h 36 min 37.33 s	+63° 19' 13.1"	4.261	5.9 ± 1.0	6.0 ± 2.1	3.3×10^8	7 × 0 at 159°	4 at 156°
1	14 h 36 min 32.46 s	+63° 20' 02.5"	...	4.3 ± 1.0	3.9 ± 1.8	...	...	...
B3J2330+3927	23 h 30 min 24.91 s	+39° 27' 11.2"	3.086	15.6 ± 1.1	22.2 ± 2.7	1.0×10^9	16 × 7 at 130°	2 at 149°
1	23 h 30 min 19.14 s	+39° 27' 03.0"	...	6.9 ± 1.1	8.2 ± 1.9	...	...	...
2	23 h 30 min 20.52 s	+39° 26' 57.7"	...	5.0 ± 1.1	3.5 ± 1.4	...	...	...
PKS1138–262	11 h 40 min 48.25 s	–26° 29' 10.1"	2.156	5.9 ± 1.1	6.7 ± 2.4	4.6×10^8	30 × 0 at 72°	11 at 86°
1	11 h 40 min 53.38 s	–26° 29' 11.9"	~2.16	9.9 ± 1.1	7.8 ± 2.2	...	17 × 0 at 60°	...
2	11 h 40 min 45.80 s	–26° 29' 56.6"	...	5.9 ± 1.1	3.1 ± 1.6	...	...	...
3	11 h 40 min 45.61 s	–26° 29' 06.6"	...	4.6 ± 1.1	2.2 ± 1.4	...	...	...
4C60.07	05 h 12 min 54.80 s	+60° 30' 51.7"	3.788	21.6 ± 1.3	23.8 ± 3.5	1.3×10^9	11 × 4 at 110°	16 at 107°
1	05 h 12 min 46.52 s	+60° 30' 35.8"	...	5.9 ± 1.3	6.3 ± 2.1	...	...	...
4C41.17	06 h 50 min 52.15 s	+41° 30' 30.8"	3.792	12.3 ± 1.2	12.0 ± 2.3	7.2×10^8	9 × 6 at 97°	20 at 48°
1	06 h 50 min 51.52 s	+41° 30' 01.5"	...	12.2 ± 1.2	21.2 ± 2.9	...	17 × 15 at 54°	...
2	06 h 50 min 49.25 s	+41° 30' 01.5"	...	7.1 ± 1.2	6.2 ± 1.9	...	8 × 0 at 50°	...
WN J0305+3525	03 h 05 min 47.42 s	+35° 25' 13.4"	3 ± 1	12.2 ± 1.0	17.6 ± 2.4	6.9×10^8	19 × 9 at 9°	2 at 64°

The coordinates of the HzRG companion sources are measured from the submillimetre maps. Values are accurate to 4–5". S_{peak} is the flux density per beam calculated at the peak of the dust emission. S_{total} is the flux density of total dust emission calculated in an aperture. A calibration uncertainty of ~10% is not included in these values. For some sources, the higher total flux densities reflect the extended nature of the dust emission, consistent with the finding that interferometric millimetre flux densities are often smaller than those calculated from single dish measurements^{23,24}. The dust mass, M_d , is calculated from the peak fluxes, assuming the following: a dust emissivity index, $\beta = 2.0$; a dust temperature, $T_d = 40$ K, typical of HzRGs; a mass absorption coefficient, κ_d (850 μm) = 0.076 m² kg^{–1}; and an $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$ cosmology. Our selection criteria result in a small spread in the calculated dust masses; an average value is $\sim 8 \times 10^8 M_\odot$. We can convert this dust mass into a conservative gas mass (molecular and atomic) assuming a standard Galactic gas-to-dust mass ratio of 200 (ref. 25), giving $1.6 \times 10^{11} M_\odot$, where $M_\odot$ is the mass of the Sun. We can confirm this estimate using the measured CO luminosities of three of the HzRGs, which can then be turned into molecular gas, and then total gas, mass using the standard Galactic conversion factor between CO luminosity and H₂ mass and a ratio of atomic-to-molecular gas of ~2. In this manner, gas masses of $(1.3\text{--}2.2) \times 10^{11} M_\odot$ have been estimated for 4C60.07, 8C1909 + 722 and B3J2330 + 3927^{23,24}. Hence the gas masses from the two methods are in reasonable agreement, and a representative estimate for the total gas mass in a HzRG with a submillimetre flux density of 10–15 mJy is thus $\sim 2 \times 10^{11} M_\odot$. While this number is arguably uncertain by a factor of a few, it is very hard to argue that it can be inflated by an order of magnitude⁵. Assuming that an order of magnitude more mass is in the form of dark matter, we estimate the halo masses to be $> 10^{12} M_\odot$. Full-width at half-maximum is from a two-dimensional gaussian fit to the 850- μm emission, deconvolved from the beam. All position angles (PA) are measured east of north. For the companion galaxies we give sizes for the brightest (highest S/N) sources only. A quoted size of zero arcseconds means that the source was not resolved in that direction. The quoted radio sizes are hotspot separations measured at 5 GHz. WJ0305 + 35 lacks a robust spectroscopic redshift, but is thought to lie at $z = 3 \pm 1$ (ref. 26).

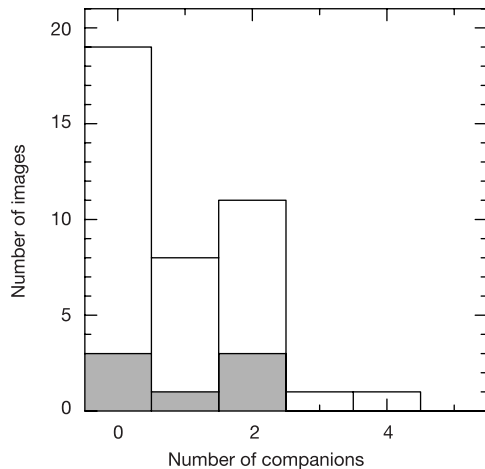


Figure 4 Comparison with hierarchical models. The predicted distribution of number of companion sources per SCUBA field (after subtracting a background of one source per map for field contamination) from our Λ -CDM simulations is shown against the observed distribution (shaded region). For the simulations we assume that the radio emission from the HzRG is powered by black holes with mass $\approx 10^9 M_\odot$ (ref. 29), and therefore these lie at the heart of dark matter haloes with masses of $\approx 10^{13} M_\odot$ (ref. 19). Allowing for intrinsic scatter and uncertainties in this estimate, we therefore searched for virialized dark matter haloes with masses in the range $12.5 < \log_{10}(M_{\text{halo}}/M_\odot) < 13.5$ at $z \approx 3$ within an N -body simulation of a region of a Λ -CDM universe, covering a co-moving volume of $(100 \text{ h}^{-1} \text{ Mpc})^3$. Parameters of the simulation were matched to the “concordance model” with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$; further details of the simulation are given elsewhere (Λ -CDM₁₀₀₀) (ref. 30). Within this simulation we identified 40 haloes which could be the counterparts of the HzRGs, and then investigated the mass distribution of other virialized dark matter haloes within a sphere of projected diameter 3' (approximately the SCUBA field of view) around each of these. Within these regions we typically found ten haloes with $M > 10^{11} M_\odot$, but on average only 1 halo with $M > 10^{12} M_\odot$.

emission and (2) the brightest submillimetre ‘companions’. The significance of these alignments is presented and quantified in Fig. 3. These results can be interpreted as follows. First, while one might be tempted to conclude that the first alignment effect is indicative of jet-induced star formation similar to the radio-ultraviolet alignment effect previously reported in HzRGs^{13,14}, the fact that in all but two cases, the radio source is also aligned with a submillimetre source on scales well beyond the size of the radio emission indicates that this is unlikely to be the complete explanation. Moreover, this model clearly cannot explain why the submillimetre emission from the HzRGs appears to be aligned with the positions of the brightest companion sources in their vicinity (in at least four out of seven cases).

There is one model that can explain both alignment effects. We propose that the brightest submillimetre companions trace the large-scale structure around the HzRG and that these directions thus contain the densest cross-sections of gas. Next, by selecting some of the very brightest known radio sources at this epoch we will then have (inadvertently, but probably inevitably) selected sources which happen to have produced jets aligned with the densest regions of gas, thereby producing very effective working surfaces and the brightest hotspots. Such a selection effect has been suggested before to explain apparent large-scale optical-radio alignment effects at lower redshift^{4,15}.

Of course, an important corollary to this interpretation is that the brightest companion sources seen in the submillimetre maps must lie at the same redshift as the radio galaxies, occupying the same large-scale structure. However, the fact that extension of this analysis to the second-brightest apparent companion object does not reveal a significant alignment effect shows that not all of the objects seen in these images need lie at the same redshift as, and be physically associated with, the radio galaxies. This is consistent with our earlier estimate of the rate of contamination in these fields by unrelated submillimetre sources.

Thus the analysis of these possible alignment effects indicates that

the submillimetre emission from the radio galaxies themselves, and that of the brightest companions, is tracing the large-scale structure around radio galaxies at $z \approx 3$. For 53W002 (a HzRG not included in our sample), it has been shown with optical spectroscopy that the brightest submillimetre companion is indeed at the same redshift as the radio galaxy¹⁶. A similar conclusion can be inferred for PKS1138–262 where the brightest submillimetre companion is coincident with one of five active galactic nuclei (AGN) forming a radio-aligned filament in the plane of the sky, and which has the same redshift as the HzRG based on narrow-band imaging of redshifted Lyman α emission¹⁷.

We have shown that these fields contain overdensities of submillimetre sources, so we now ask what is the typical mass of these galaxies. We can do this by assuming that they lie at the same redshift as the HzRGs. The dust masses of the HzRGs, calculated from their 850- μm fluxes¹⁸, are given in Table 1. Using standard assumptions we can convert these dust masses into total gas masses, and hence estimate the masses of their associated dark matter haloes (see legend to Table 1 for details). We estimate that they reside in dark matter haloes with masses in excess of $10^{12}M_{\odot}$. Because the submillimetre flux densities of the companions are similar to those of the HzRGs, we can infer that they also reside in dark matter haloes with masses in excess of $10^{12}M_{\odot}$. We therefore conclude that these regions can contain more than one very massive and gas-rich galaxy of $>10^{12}M_{\odot}$. Can we reconcile this result with the hierarchical models of galaxy and structure formation?

By design, we have selected highly biased regions of the high-redshift Universe by imaging around some of the most luminous radio galaxies at these epochs, which are expected to host some of the most massive black holes. We can attempt to quantify the impact of this bias by exploring the predictions of numerical simulations of the collapse of Cold Dark Matter (CDM) haloes within a Λ -dominated universe. The details and results of these simulations are given in Fig. 4. These simulations do indeed appear to be in accord with the observations, that is, both data and theory suggest that the submillimetre companions revealed by our SCUBA imaging are associated with dark-matter haloes more massive than $10^{12}M_{\odot}$. This conclusion is, however, subject to an assumption about the duty cycles of the submillimetre luminous phase; if true, it would imply that the vast majority of such massive haloes are actively forming stars at the epochs sampled by these images. This is not unreasonable, especially since we have deliberately targeted fields in which at least one massive object (the HzRG) is actively engaged in intense star formation.

Pursuing this comparison one step further, we can explore the predicted and observed distribution of companion objects in this high-mass range. In fact, although the average number of such companions in the simulated SCUBA images centred on the HzRG is one, this average arises from a skewed distribution. As shown in Fig. 4, of the 40 high-mass haloes investigated, 50% have no such massive companions, with the average of one resulting from the fact that, of the remaining half, 40% have two companions or more. We have over-plotted in Fig. 4 the corresponding histogram of companion incidence for the seven HzRGs imaged in this study (after statistically correcting for field contamination). This comparison at least illustrates that the distribution of companion incidence in our data is consistent with the prediction of the simulations, that is, typically we see either no companion, or two or more companions.

The average number of companions, the distribution of companion incidence, and the inferred baryonic gas masses of the companions are all most consistent with the interpretation that the SCUBA sources uncovered in this study are the progenitors of massive present-day cluster ellipticals. In this case, because elliptical galaxies are known to contain massive black holes with mass proportional to that of the spheroid¹⁹, it is reasonable to assume that the black hole and stellar mass grow coevally from the same gas

reservoir. This in turn suggests that the companion objects should contain buried AGN^{20,21}. The first results show a high rate of correspondence between the submillimetre companions and luminous X-ray sources^{16,22}, suggesting that this is indeed the case. We are now pursuing high-resolution follow-up observations of these fields at millimetre and optical/near-infrared wavelengths. The former will yield information on source structure and the latter will reveal counterparts for spectroscopic redshift determination on 10-m class telescopes. $\square$

Received 7 April; accepted 5 August 2003; doi:10.1038/nature01976.

- Ellis, R. S. *et al.* The homogeneity of spheroidal populations in distant clusters. *Astrophys. J.* **483**, 582–596 (1997).
- Dunlop, J. S. *et al.* Detection of a large mass of dust in a radio galaxy at redshift $z = 3.8$. *Nature* **370**, 347–349 (1994).
- Kauffmann, G. The age of elliptical galaxies and bulges in a merger model. *Mon. Not. R. Astron. Soc.* **281**, 487–492 (1996).
- West, M. J. Anisotropic mergers at high redshifts: the formation of cD galaxies and powerful radio sources. *Mon. Not. R. Astron. Soc.* **268**, 79–102 (1994).
- Hughes, D. H., Dunlop, J. S. & Rawlings, S. High-redshift radio galaxies and quasars at submillimetre wavelengths: assessing their evolutionary status. *Mon. Not. R. Astron. Soc.* **289**, 766–782 (1997).
- Archibald, E. N. *et al.* A submillimetre survey of the star-formation history of radio galaxies. *Mon. Not. R. Astron. Soc.* **323**, 417–444 (2001).
- Reuland, M., Röttgering, H. & van Breugel, W. in *Radio Galaxies: Past, Present and Future* (Elsevier, in the press); preprint at (<http://arXiv.org/astro-ph/0303321>) (2003).
- Holland, W. S. *et al.* SCUBA: a common-user submillimetre camera operating on the James Clerk Maxwell Telescope. *Mon. Not. R. Astron. Soc.* **303**, 659–672 (1999).
- Graham, A., Lauer, T. R., Colless, M. & Postman, M. Brightest cluster galaxy profile shapes. *Astrophys. J.* **465**, 534–547 (1996).
- Mihos, J. C. & Hernquist, L. Gasdynamics and starbursts in major mergers. *Astrophys. J.* **464**, 641–663 (1996).
- Smail, I., Ivison, R. J., Blain, A. W. & Kneib, J.-P. The nature of faint submillimetre selected galaxies. *Mon. Not. R. Astron. Soc.* **331**, 495–520 (2002).
- Scott, S. E. *et al.* The SCUBA 8-m Jy survey I. Submillimetre maps, sources and number counts. *Mon. Not. R. Astron. Soc.* **331**, 817–838 (2002).
- Chambers, K. C., Miley, G. K. & van Breugel, W. Alignment of radio and optical orientations in high-redshift radio galaxies. *Nature* **329**, 604–606 (1987).
- McCarthy, P. J., van Breugel, W., Spinrad, H. & Djorgovski, S. A correlation between the radio and optical morphologies of distant 3CR radio galaxies. *Astrophys. J.* **321**, L29–L33 (1987).
- Eales, S. A. A new theory for the alignment effect. *Astrophys. J.* **397**, 49–54 (1992).
- Smail, I. *et al.* A SCUBA galaxy in the protocluster around 52W002 at $z = 2.4$. *Astrophys. J.* **583**, 551–558 (2003).
- Pentericci, L. *et al.* A Chandra study of X-ray sources in the field of the $z = 2.16$ radio galaxy MRC 1138–262. *Astron. Astrophys.* **396**, 109–115 (2002).
- Hildebrand, R. H. The determination of cloud masses and dust characteristics from submillimetre thermal emission. *Q. J. R. Astron. Soc.* **24**, 267–282 (1983).
- Magorrian, J. *et al.* The demography of massive dark objects in galactic centers. *Astron. J.* **115**, 2285–2305 (1998).
- Kauffmann, G. & Haehnelt, M. A unified model for the evolution of galaxies and quasars. *Mon. Not. R. Astron. Soc.* **311**, 576–588 (2000).
- Page, M. J., Stevens, J. A., Mittaz, J. P. D. & Carrera, F. J. Submillimetre evidence for the coeval growth of massive black holes and galaxy bulges. *Science* **294**, 2516–2518 (2001).
- Smail, I. *et al.* Chandra detections of SCUBA galaxies around high- z radio sources. *Astrophys. J.* (in the press).
- De Breuck, C. *et al.* CO emission and associated HI absorption from a massive gas reservoir surrounding the $z = 3$ radio galaxy B3 J2330+393. *Astron. Astrophys.* **401**, 911–925 (2003).
- Papadopoulos, P. P. *et al.* CO(4–3) and dust emission in two powerful high- z radio galaxies, and CO lines at high redshifts. *Astrophys. J.* **528**, 626–636 (2000).
- Gorenstein, P. Empirical relation between interstellar X-ray absorption and optical extinction. *Astrophys. J.* **198**, 95–101 (1975).
- Reuland, M. *et al.* An obscured radio galaxy at high redshift. *Astrophys. J.* **582**, L71–L74 (2003).
- Ivison, R. J. *et al.* An excess of submillimetre sources near 4C41.17: a candidate proto-cluster at $z = 3.8$? *Astrophys. J.* **542**, 27–34 (2000).
- Omont, A. *et al.* Molecular gas and dust around a radio-quiet quasar at redshift 4.69. *Nature* **382**, 428–431 (1996).
- Dunlop, J. S. *et al.* Quasars, their host galaxies, and their central black holes. *Mon. Not. R. Astron. Soc.* **340**, 1095–1135 (2003).
- Percival, W. J., Scott, D., Peacock, J. A. & Dunlop, J. S. The clustering of halo mergers. *Mon. Not. R. Astron. Soc.* **338**, L31–L35 (2003).

Acknowledgements The JCMT is run by the Joint Astronomy Centre on behalf of the Particle Physics and Astronomy Research Council (PPARC). J.A.S., J.S.D., I.R.S. and W.J.P. acknowledge funding from PPARC, the Royal Society and the Leverhulme Trust. The work of W.J.M.v.B. and M.R. was performed under the auspices of the US Department of Energy, National Nuclear Security Administration by the University of California, Lawrence Livermore National Laboratory.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.S. (jas@roe.ac.uk).

Experimental realization of a one-atom laser in the regime of strong coupling

J. McKeever, A. Boca, A. D. Boozer, J. R. Buck & H. J. Kimble

Norman Bridge Laboratory of Physics 12-33, California Institute of Technology, Pasadena, California 91125, USA

Conventional lasers (from table-top systems to microscopic devices) typically operate in the so-called weak-coupling regime, involving large numbers of atoms and photons; individual quanta have a negligible impact on the system dynamics. However, this is no longer the case when the system approaches the regime of strong coupling for which the number of atoms and photons can become quite small. Indeed, the lasing properties of a single atom in a resonant cavity have been extensively investigated theoretically^{1–11}. Here we report the experimental realization of a one-atom laser operated in the regime of strong coupling. We exploit recent advances¹² in cavity quantum electrodynamics that allow one atom to be isolated in an optical cavity in a regime for which one photon is sufficient to saturate the atomic transition. The observed characteristics of the atom–cavity system are qualitatively different from those of the familiar many-atom case. Specifically, our measurements of the intracavity photon number versus pump intensity indicate that there is no threshold for lasing, and we infer that the output flux from the cavity mode exceeds that from atomic fluorescence by more than tenfold. Observations of the second-order intensity correlation function demonstrate that our one-atom laser generates manifestly quantum (nonclassical) light, typified by photon antibunching and sub-poissonian photon statistics.

The usual laser theories rely on system-size expansions in inverse powers of critical atom and photon numbers ($N_0, n_0 \gg 1$), and arrive at a consistent form for the laser characteristics^{13–17}. By contrast, over the past twenty years, technical advances on various fronts have pushed laser operation to regimes of ever smaller atom and photon number, pressing toward the limit of ‘strong coupling’ for which ($N_0, n_0 \ll 1$ (ref. 18)). Significant milestones include the realization of one- and two-photon micromasers^{19–21}, as well as microlasers in atomic and condensed matter systems^{22–24}.

As illustrated in Fig. 1, our experiment consists of a single caesium atom trapped in a far-off-resonance trap (FORT) within a high-finesse optical cavity^{12,25}. The lasing transition $6P_{3/2}, F' = 3' \rightarrow 6S_{1/2}, F = 4$ is nearly resonant with and strongly coupled to a single mode of this cavity. The coupling is parameterized by the Rabi frequency $2g_0$ for a single quantum of excitation, and the atom and field have amplitude decay rates γ and κ , respectively. The upper level $F' = 3'$ is pumped by the external drive Ω_3 , while effective decay of the lower level $F = 4$ takes place via the combination of the drive Ω_4 and decay γ_{34} , $4 \rightarrow 4' \rightarrow 3$. In essential character this system is analogous to a Raman scheme with pumping $3 \rightarrow 3'$, lasing $3' \rightarrow 4$, and decay $4 \rightarrow 3$. Of particular relevance to our work are detailed treatments of the ion-trap laser^{6–9}.

We emphasize that a ‘one-and-the-same’ atom laser as illustrated in Fig. 1 is quite distinct from ‘single-atom’ micromasers^{19–21} and lasers²² for which steady state is reached through the incremental contributions of many atoms that transit the cavity, even if one by one^{19,20} or few by few²². By contrast, in our experiment steady state is reached with one-and-the-same atom over a time interval $\delta t \approx 10^{-7}$ s that is much shorter than the trap lifetime $\Delta t \approx 0.05$ s. Our pumped atom–cavity system provides a continuous source of nonclassical light as a gaussian beam for the entire duration that an atom is trapped.

Because conventional lasers operate in the limit ($N_0, n_0 \gg 1$), there is a generic form associated with the laser threshold in the transition from nonlasing to lasing action that is independent of the model system^{13,26}. However, as the system size is reduced, the sharpness of the laser ‘turn on’ is lost, with then no clear consensus about how to define the lasing threshold²⁶. Well into the regime of strong coupling ($N_0, n_0 \ll 1$), even the familiar qualitative characteristics of a laser (for example, the statistical properties of the output light) are profoundly altered, leaving open the question of how to recognize a laser in this new regime.

To address this question, we have carried out extensive theoretical analyses for a four-state model based upon Fig. 1b for parameters relevant to our experiment. A synopsis of relevant results from this work is given in the Supplementary Information, with the full treatment presented in ref. 27. In brief, the steady-state solutions obtained from a semiclassical theory exhibit familiar characteristics of conventional lasers, including a clearly defined laser threshold and population inversion. The condition $C_1 \gg 1$ is required to observe threshold behaviour for one atom pumped inside the resonator, where for our experiment the cooperativity parameter $C_1 = 1/N_0 \approx 12$. By contrast, the fully quantum analysis for the four-state model results in qualitatively different characteristics. In particular, the input–output relationship for the mean intracavity photon number $\bar{n}$ versus the pump intensity $I_3 = (\Omega_3/2\gamma)^2$ has several key features to be compared with experimental results presented below, namely the immediate onset of emission (‘thresholdless’ behaviour), and the saturation and eventual quenching of the output.

Our actual experiment is more complex than indicated by the simple drawing in Fig. 1, with many of the technical aspects described in more detail in refs 12 and 25. In brief, the principal cavity QED (cQED) parameters of our system are $g_0/2\pi = 16$ MHz, $\kappa/2\pi = 4.2$ MHz, and $\gamma/2\pi = 2.6$ MHz, where g_0 is based upon the reduced dipole moment for the $6S_{1/2}, F = 4 \leftrightarrow 6P_{3/2}, F' = 3'$ transition in atomic caesium. Strong coupling is thereby achieved ($g_0 \gg (\kappa, \gamma)$), resulting in critical photon and atom numbers $n_0 \equiv \gamma^2/(2g_0^2) \approx 0.013$, $N_0 \equiv 2\kappa\gamma/g_0^2 \approx 0.084$.

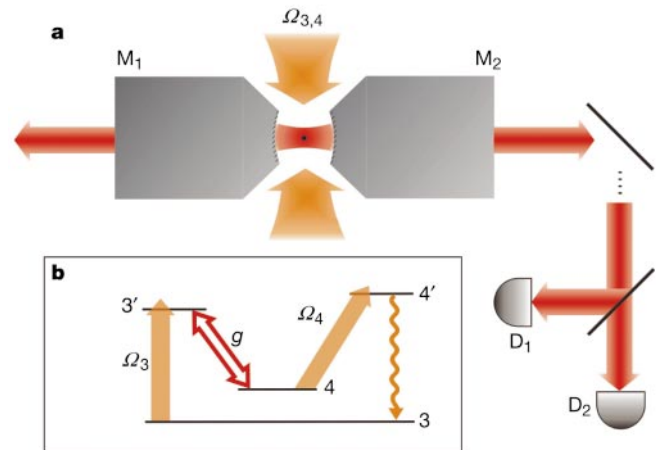


Figure 1 A simplified schematic of the experiment. **a**, A caesium atom (black dot) is trapped inside a high-finesse optical cavity formed by the curved, reflective surfaces of mirrors M_1 and M_2 . Light generated by the atom's interaction with the resonant cavity mode propagates as a gaussian beam to single-photon detectors D_1 and D_2 . **b**, The relevant transitions involve the $6S_{1/2}, F = 3, 4 \leftrightarrow 6P_{3/2}, F' = 3', 4'$ levels of the D_2 line at 852.4 nm in atomic caesium. Strong coupling at rate g is achieved for the lasing transition $F' = 3' \rightarrow F = 4$ near a cavity resonance. Pumping of the upper level $F' = 3'$ is provided by the field Ω_3 , while recycling of the lower level $F = 4$ is achieved by way of the field Ω_4 ($4 \rightarrow 4'$) and spontaneous decay back to $F = 3$. Decay $(3', 4') \rightarrow (3, 4)$ is also included in our model. Relevant cavity parameters are length $l_0 = 42.2 \mu\text{m}$, waist $w_0 = 23.6 \mu\text{m}$, and finesse $\mathcal{F} = 4.2 \times 10^5$ at $\lambda_{D_2} = 852 \text{ nm}$.

Atoms are trapped in the cavity by means of a FORT²⁸ with wavelength $\lambda_F = 935.6$ nm, which is matched to a TEM₀₀ mode along the cavity axis. For all experiments herein, the trap depth is $U_0/k_B = 2.3$ mK (47 MHz) where k_B is the Boltzmann constant. The FORT has the important feature that the potential for the atomic centre-of-mass motion is only weakly dependent on the atom's internal state¹².

After the trap-loading stage (as described in the section on Methods), the transverse $\Omega_{3,4}$ fields are switched to pump and recycle the atomic population in the fashion depicted in Fig. 1b. Two examples of the resulting output counts versus time are shown in Fig. 2. By averaging traces such as these, we arrive at an average signal level versus time, as shown in the inset to Fig. 2a. Typical lifetimes for a trapped atom in the presence of the driving $\Omega_{3,4}$ fields are 50–100 ms, which should be compared to the lifetimes of 2–3 s recorded in the absence of these fields¹². Significantly, the approximately exponential decay of the signal with time does not result from a time-dependent diminution of the flux from single trapped atoms, but rather from the average of many events each of a variable duration. That is, for a given set of external control parameters, each atom gives a reasonably well-defined output flux over the time that it is trapped.

For a fixed set of operating conditions, we collect a set of 60–300 traces as in Fig. 2, determine the average output flux for each trace, and find the mean and variance, as well as the trap lifetime for the set. Figure 3 displays a collection of such measurements for the mean intracavity photon number $\bar{n}$ as a function of the dimensionless pump intensity x , scaled in units of the fixed recycling intensity (see section on Methods). More precisely, the parameter x is the ratio of measured intensities, and can be written as $x \equiv (7/9)(I_3/I_4)$, where $I_{3,4} \equiv (\Omega_{3,4}/2\gamma)^2$. The factor of (7/9) is needed because the

two transitions have different dipole moments. For these measurements, we estimate that the incoherent sum of intensities of the four Ω_4 beams is about 50 mW cm^{-2} , which corresponds to $I_4 \approx 13$. The output count rate at detectors $D_{1,2}$ is converted to intracavity photon number using the known propagation and detection efficiency $\xi = 0.05$.

Important features of the data shown in Fig. 3 include the prompt onset of output flux $\kappa\bar{n}$ emerging through the cavity mirrors $M_{1,2}$ as the pump intensity I_3 is increased from zero. In a regime of strong coupling, the atom–cavity system behaves as a ‘thresholdless’ device. With further increases in pump intensity I_3 , the output flux saturates at a maximum value $\kappa\bar{n}_{\text{max}}$ around $x \approx 0.1$. We attribute this behaviour to a bottleneck associated with the recycling of population $4 \rightarrow 4' \rightarrow 3$, with the rate-limiting step in the recycling process being spontaneous decay $4' \rightarrow 3$ at rate γ_{34} in the limit of large Rabi frequency $\Omega_4 \gg \gamma$. For a single intracavity atom, quanta can be deposited into the cavity mode no faster than the maximum recycling rate. As the pump level I_3 is increased beyond $x \approx 1$, the output flux $\kappa\bar{n}$ gradually drops, presumably owing to splitting of the pumped excited state $F' = 3'$ by the Autler–Townes effect, although this is still under investigation. Heating of the atomic motion at higher pump levels is certainly a concern as well; however, our simulations, which do not incorporate atomic motion, show the same trend as in Fig. 3 (ref. 27).

Beyond these considerations, we have also undertaken extensive theoretical analyses based both upon the four-state model shown in Fig. 1, as well as on the full set of Zeeman states for each of the levels $F = 3, 4$ and $F' = 3', 4'$ and two cavity modes, one for each of two orthogonal polarizations²⁷. These analyses are in reasonable accord with the principal features of the data in Fig. 3. Moreover, our quantum simulations support the conclusion that the range of coupling values g that contribute to our results is restricted roughly to $0.5g_0 \leq g \leq g_0$. Furthermore, the simulations yield information about the atomic populations, from which we deduce that the rate of emission from the cavity $\kappa\bar{n}$ exceeds that by way of fluorescent decay $3' \rightarrow 4$, $\gamma_{43'} \langle \sigma_{3',3'} \rangle$, by roughly tenfold over the range of pump intensity I_3 shown in Fig. 3, where $\langle \sigma_{3',3'} \rangle$ is the steady-state population in level $3'$.

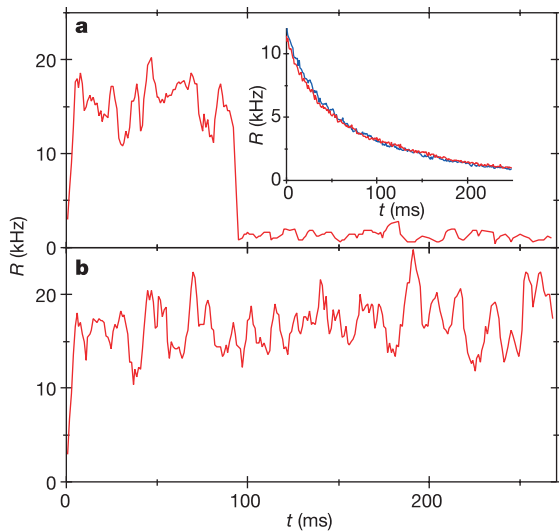


Figure 2 Total counting rate R recorded by detectors $D_{1,2}$ is displayed as a function of time for two separate trapped atoms, with the counts summed over 5-ms bins. At $t = 0$, the $\Omega_{3,4}$ fields are switched to predetermined values of intensity and detuning. In **a**, the atom is trapped for $t \approx 90$ ms before escaping, with the background level due to scattered light from the $\Omega_{3,4}$ fields and detector dark counts evident as the residual output at later times. In **b**, the atom (atypically) remains trapped for the entire observation cycle ≈ 270 ms and then is dumped. The inset in **a** displays R versus time obtained by averaging 400 such traces. Two cases are shown; in one, the number of atoms delivered to the cavity mode has been diminished by about two fold. The curves are nearly identical, so we conclude that cases with $N > 1$ atom play a negligible role. The overall detection efficiency $\xi = 0.05$ from intracavity photon to a detection event at D_1 or D_2 is made up of the following factors: $\eta = 0.60$ cavity escape efficiency, $T = 0.50$ for only mirror M_2 output, $\zeta = 0.33$ propagation efficiency from M_2 to $D_{1,2}$, and $\alpha = 0.5$ detection quantum efficiency at $D_{1,2}$.

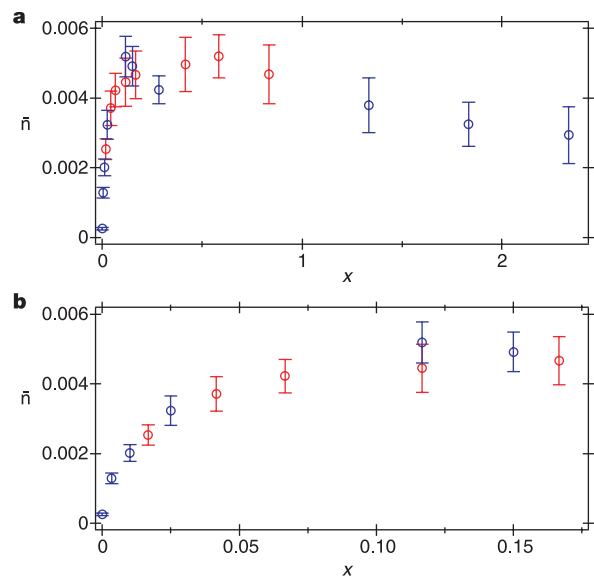


Figure 3 The intracavity photon number $\bar{n} \pm \sigma_n$ inferred from measurements as in Fig. 2, is plotted as a function of dimensionless pump intensity $x \equiv (7/9)(I_3/I_4)$ for fixed $I_4 = 13$ over two ranges of pump level x . **a**, $\bar{n}$ versus x is shown over the entire range $x = 0$ to 2.33 recorded in our measurements. **b**, An expanded scale displays $\bar{n}$ for small x . The immediate onset of emission supports the conclusion of ‘thresholdless’ lasing. The two independent sets of measurements (red and blue points) agree reasonably well.

To investigate the quantum-statistical characteristics of the light emerging in the TEM₀₀ mode of the cavity output, we probe the photon statistics of the light by way of the two single-photon detectors D_{1,2} illustrated in Fig. 1. From the cross-correlation of the resulting binned photon arrival times and the mean counting rates of the signals and the background, we construct the normalized intensity correlation function (see the Supplementary Information)

$$g^{(2)}(\tau) = \frac{\langle : \hat{I}(t) \hat{I}(t+\tau) : \rangle}{\langle : \hat{I}(t) : \rangle^2} \quad (1)$$

where the colons denote normal and time ordering for the intensity operators $\hat{I}$ (ref. 15). Over the duration of the trapping events, we find no evidence that $\langle : \hat{I}(t) : \rangle$ is a function of t , although we do not have sufficient data to confirm quantitatively stationarity of the underlying processes.

Examples of two measurements for $g^{(2)}(\tau)$ are given in Fig. 4. In Fig. 4 a and b, we again have $I_4 \approx 13$ and the pump intensity I_3 is set for operation with $x \approx 0.83$ well beyond the 'knee' in $\bar{n}$ versus x , while in Fig. 4 c and d, the pump level is decreased to $x \approx 0.17$ near the peak in $\bar{n}$. Significantly, in each case these measurements demonstrate that the light from the atom-cavity system is manifestly quantum (that is, nonclassical) and exhibits photon antibunching $g^{(2)}(0) < g^{(2)}(\tau)$ and sub-poissonian photon statistics $g^{(2)}(0) < 1$ (ref. 15). The actual coincidence data $n(\tau)$ used to obtain $g^{(2)}(\tau)$ are presented in the Supplementary Information. Significantly, these data directly provide evidence of the nonclassical character of the emitted light, with relatively minor corrections for background light required for the determination of $g^{(2)}(\tau)$.

Beyond the nonclassical features around $\tau \approx 0$, $g^{(2)}(\tau)$ also exhibits excess fluctuations extending over $\tau \approx \pm 1 \mu\text{s}$, with $g_{\text{max}}^{(2)}(\tau) \approx 1.7$. Fluctuations in the intensity of the intracavity light over these timescales are presumably related to the stochastic character of the pumping $3 \rightarrow 3'$ and recycling $4 \rightarrow 4' \rightarrow 3$ processes for a single, multi-state atom. Also of significance is the interplay of atomic motion and optical pumping into dark states by the $\Omega_{3,4}$ fields (which is responsible for cooling; ref. 29 and references therein), as well as Larmor precession that arises from residual ellipticity in polarization of the intracavity FORT^{12,30}. Indeed, in Fig. 4a, c there is a hint of an oscillatory variation in $g^{(2)}(\tau)$ with period $\tau \approx \pm 2 \mu\text{s}$. Fourier transformation of the associated coincidence data leads to a small peak at about 500 kHz, which is near to the predicted frequency for axial motion

of a trapped caesium atom at the bottom of the FORT potential, as well as to the Larmor frequency inferred from other measurements.

In agreement with the trend predicted by the four-state model discussed in the Supplementary Information, $g^{(2)}(0)$ increases with increasing pump intensity, with a concomitant decrease in these nonclassical effects. Moreover, our experimental observations of $g^{(2)}(\tau)$ are described reasonably well by the results obtained from more detailed quantum simulations based upon the entire manifold of Zeeman states for the caesium atom, two cavity modes with orthogonal polarizations, and a simple model to describe the polarization gradients of the $\Omega_{3,4}$ fields²⁷.

The realization of this strongly coupled one-atom laser is significant on several fronts. From the perspective of the dynamics of open quantum systems, our system demonstrates the radical departures from conventional laser operation wrought by strong coupling for the quantized light-matter interaction. On a more practical level, throughout the interval when an atom is trapped (which is determined in real time), our system provides an approximately stationary source of nonclassical light in a collimated, gaussian beam, as has been anticipated in the literature on one-atom lasers^{1,3-6,8-11}, and which has diverse applications. Some remaining technical issues in our work are to improve the modelling and measurements related to atomic motion, both within the FORT potential and through the polarization gradients of the $\Omega_{3,4}$ fields. We have employed our quantum simulations to calculate the optical spectrum of the light output, and have devised a scheme for its measurement. □

Methods

While the atom is trapped in a standing-wave FORT along the cavity axis, another set of fields (designated by $\Omega_{3,4}$ in Fig. 1) propagate in the plane transverse to the cavity axis and illuminate the region between the cavity mirrors. These fields are used not only for the pumping scheme described in association with the operation of the one-atom laser with strong coupling, but also for cooling in the trap-loading phase. Each $\Omega_{3,4}$ field consists of two orthogonal pairs of counter-propagating beams in a $\sigma^+ - \sigma^-$ configuration.

Unfortunately, it is difficult to calibrate accurately the intensities $I_{3,4}$ for the $\Omega_{3,4}$ beams at the location of the atom in the region between the cavity mirrors. We estimate that our knowledge of either intensity is uncertain by an overall scale factor of about 2. However, we do know the ratio of intensities much more accurately than either intensity individually, and therefore plot the data in Fig. 3 as a function of this ratio.

In the pumping stage of the experiment, the fields are tuned 10 MHz blue of $F = 3 \rightarrow F' = 3'$ in the case of the Ω_3 beams and 17 MHz blue of $F = 4 \rightarrow F' = 4'$ in the case of the Ω_4 fields. The detuning between the $3' \rightarrow 4$ transition at $\omega_{4,3}$ and the cavity resonance ω_C is $\Delta_{CA} = \omega_C - \omega_{4,3} = 2\pi \times 9 \text{ MHz}$. These detunings are chosen operationally in a trade-off between achieving a large cavity output flux from the $3' \rightarrow 4$ transition while maintaining a reasonable lifetime for the trapped atom despite heating from the various fields²⁹. The cavity length itself is actively stabilized with an auxiliary laser at wavelength $\lambda_C = 835.8 \text{ nm}$ that does not interfere with the trapping or the cQED interactions.

Our experimental protocol begins with the formation of a magneto-optical trap (MOT) above the cavity. After a stage of sub-Doppler cooling, the cloud of atoms is released. The $\Omega_{3,4}$ beams are then used as cooling beams (with independent settings of intensity and detuning) to load an atom into the FORT¹². About ten atoms transit the cavity mode after each MOT drop, and the loading efficiency is set such that an atom is loaded into the FORT once every 3–10 drops. We then switch the intensities and detunings of the transverse fields $\Omega_{3,4}$ to the pumping configuration and record the cavity output by way of the single-photon detectors D_{1,2} shown in Fig. 1. Each photoelectric pulse from D_{1,2} is stamped with its time of detection (1-ns resolution) and then stored for later analysis, with examples of the record of output counts versus time displayed in Fig. 2.

Received 9 July; accepted 8 August 2003; doi:10.1038/nature01974.

- Mu, Y. & Savage, C. M. One-atom lasers. *Phys. Rev. A* **46**, 5944–5954 (1992).
- Ginzel, C., Briegel, H.-J., Martini, U., Englert, B.-G. & Schenzle, A. Quantum optical master equations: The one-atom laser. *Phys. Rev. A* **48**, 732–738 (1993).
- Pellizzari, T. & Ritsch, H. Preparation of stationary Fock states in a one-atom Raman laser. *Phys. Rev. Lett.* **72**, 3973–3976 (1994).
- Pellizzari, T. & Ritsch, H. Photon statistics of the three-level one-atom laser. *J. Mod. Opt.* **41**, 609–623 (1994).
- Horak, P., Gheri, K. M. & Ritsch, H. Quantum dynamics of a single-atom cascade laser. *Phys. Rev. A* **51**, 3257–3266 (1995).
- Meyer, G. M., Briegel, H.-J. & Walther, H. Ion-trap laser. *Europhys. Lett.* **37**, 317–322 (1997).
- Löffler, M., Meyer, G. M. & Walther, H. Spectral properties of the one-atom laser. *Phys. Rev. A* **55**, 3923–3930 (1997).
- Meyer, G. M., Löffler, M. & Walther, H. Spectrum of the ion-trap laser. *Phys. Rev. A* **56**, R1099–R1102 (1997).
- Meyer, G. M. & Briegel, H.-J. Pump-operator treatment of the ion-trap laser. *Phys. Rev. A* **58**, 3210–3220 (1998).

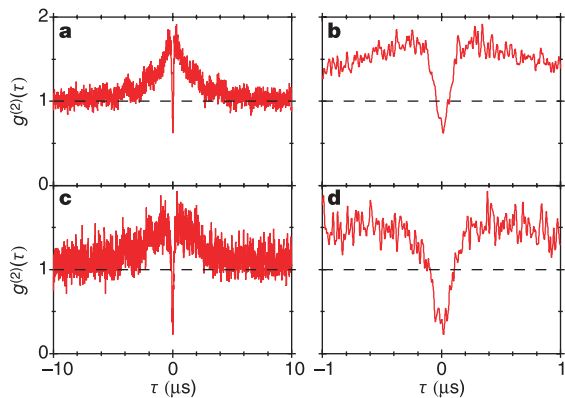


Figure 4 The intensity correlation function $g^{(2)}(\tau)$ is given for two values of the pump intensity in **a–d**, where **b** and **d** show the central features of **a** and **c** over a smaller range of τ . Panels **a** and **b** are for pump intensity parameter $x = 0.83$, whereas for **c** and **d**, $x = 0.17$. Note that the light exhibits sub-poissonian photon statistics and antibunching. All traces have been 'smoothed' by convolution with a gaussian function of width $\sigma = 5 \text{ ns}$.

10. Jones, B., Ghose, S., Clemens, J. P., Rice, P. R. & Pedrotti, L. M. Photon statistics of a single atom laser. *Phys. Rev. A* **60**, 3267–3275 (1999).
11. Kilin, S. Ya. & Karlovich, T. B. Single-atom laser: Coherent and nonclassical effects in the regime of a strong atom-field correlation. *JETP* **95**, 805–819 (2002).
12. McKeever, J. *et al.* State-insensitive cooling and trapping of single atoms in an optical cavity. *Phys. Rev. Lett.* **90**, 133602 (2003).
13. Sargent, M. III, Scully, M. O. & Lamb, W. E. Jr *Laser Physics* (Addison-Wesley, Reading, MA, 1974).
14. Haken, H. *Laser Theory* (Springer, Berlin, 1984).
15. Mandel, L. & Wolf, E. *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, Cambridge, 1995).
16. Carmichael, H. J. *Statistical Methods in Quantum Optics I* (Springer, Berlin, 1999).
17. Gardiner, C. W. & Zoller, P. *Quantum Noise* (Springer, Berlin, 2000).
18. Kimble, H. J. Strong interactions of single atoms and photons in cavity QED. *Phys. Scr.* **T76**, 127–137 (1998).
19. Raithel, G., Wagner, C., Walther, H., Narducci, L. M. & Scully, M. O. *Cavity Quantum Electrodynamics* (ed. Berman, P.) 57–121 (Academic, San Diego, 1994).
20. Haroche, S. & Raimond, J. M. *Cavity Quantum Electrodynamics* (ed. Berman, P.) 123–170 (Academic, San Diego, 1994).
21. Meystre, P. in *Progress in Optics* Vol. XXX (ed. Wolf, E.) 261–355 (Elsevier, Amsterdam, 1992).
22. An, K. & Feld, M. S. Semiclassical four-level single-atom laser. *Phys. Rev. A* **56**, 1662–1665 (1997).
23. Chang, R. K. & Campillo, A. J. (eds) *Optical Processes in Microcavities* (World Scientific, Singapore, 1996).
24. Vahala, K. J. Optical microcavities. *Nature* **424**, 839–846 (2003).
25. Ye, J., Vernooij, D. W. & Kimble, H. J. Trapping of single atoms in cavity QED. *Phys. Rev. Lett.* **83**, 4987–4990 (1999).
26. Rice, P. R. & Carmichael, H. J. Photon statistics of a cavity-QED laser: A comment on the laser-phase-transition analogy. *Phys. Rev. A* **50**, 4318–4329 (1994).
27. Boozer, A. D., Boca, A., Buck, J. R., McKeever, J. & Kimble, H. J. Comparison of theory and experiment for a one-atom laser in a regime of strong coupling. *Phys. Rev. A* (submitted); preprint available at (<http://lanl.arxiv.org/archive/quant-ph>).
28. Metcalf, H. J. & van der Straten, P. *Laser Cooling and Trapping* (Springer, New York, 1999).
29. Boiron, D. *et al.* Laser cooling of cesium atoms in gray optical molasses down to 1.1 μ K. *Phys. Rev. A* **53**, R3734–R3737 (1996).
30. Corwin, K. L., Kuppens, S. J. M., Cho, D. & Wieman, C. E. Spin-polarized atoms in a circularly polarized optical dipole trap. *Phys. Rev. Lett.* **83**, 1311–1314 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We gratefully acknowledge interactions with K. Birnbaum, C.-W. Chou, A. C. Doherty, L.-M. Duan, T. Lynn, T. Northup, S. Polyakov and D. M. Stamper-Kurn. This work was supported by the National Science Foundation, by the Caltech MURI Center for Quantum Networks, and by the Office of Naval Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.J.K. (hjkimble@caltech.edu).

Quantum critical behaviour in a high- T_c superconductor

D. van der Marel^{1*}, H. J. A. Molegraaf^{1*}, J. Zaanen², Z. Nussinov^{2*}, F. Carbone^{1*}, A. Damascelli^{3*}, H. Eisaki^{3*}, M. Greven³, P. H. Kes² & M. Li²

¹Materials Science Centre, University of Groningen, 9747 AG Groningen, The Netherlands

²Leiden Institute of Physics, Leiden University, 2300 RA Leiden, The Netherlands

³Department of Applied Physics and Stanford Synchrotron Radiation Laboratory, Stanford University, California 94305, USA

* Present addresses: Département de Physique de la Matière Condensée, Université de Genève, CH-1211 Genève 4, Switzerland (D.v.d.M., H.J.A.M., F.C.); Los Alamos National Laboratories, Los Alamos, New Mexico 87545, USA (Z.N.); Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, V6T 1Z1, Canada (A.D.); Low-Temperature Physics Group, National Institute of Advanced Industrial Science and Technology, Umezono, Tsukuba, 305-8568, Japan (H.E.)

Quantum criticality is associated with a system composed of a nearly infinite number of interacting quantum degrees of freedom at zero temperature, and it implies that the system looks on average the same regardless of the time- and length scale on which it is observed. Electrons on the atomic scale do not exhibit such symmetry, which can only be generated as a collective phenomenon through the interactions between a

large number of electrons. In materials with strong electron correlations a quantum phase transition at zero temperature can occur, and a quantum critical state has been predicted^{1,2}, which manifests itself through universal power-law behaviours of the response functions. Candidates have been found both in heavy-fermion systems³ and in the high-transition temperature (high- T_c) copper oxide superconductors⁴, but the reality and the physical nature of such a phase transition are still debated^{5–7}. Here we report a universal behaviour that is characteristic of the quantum critical region. We demonstrate that the experimentally measured phase angle agrees precisely with the exponent of the optical conductivity. This points towards a quantum phase transition of an unconventional kind in the high- T_c superconductors.

In the quantum theory of collective fields one anticipates order at small coupling constant, and for increasing coupling one expects at some point a phase transition to a quantum-disordered state. Quantum criticality in the copper oxides, if it exists, occurs as a function of charge carrier doping x , at a particular doping level x_c close to where the superconducting phase transition temperature reaches its maximum value. When this phase transition is continuous, a critical state is realized right at the transition, which is characterized by scale invariance resulting in the above-mentioned power-law response up to some (non-universal) high-energy cut-off Ω .

The optical conductivity, $\sigma(\omega) = \sigma_1(\omega) + i\sigma_2(\omega)$, is the absorptive (σ_1) and reactive (σ_2) current response to a time-varying external electrical field of frequency ω , and is usually expressed as the correlation function of the currents $\mathbf{j}(\tau_1)$ and $\mathbf{j}(\tau_2)$ at times τ_1 and τ_2 , which is $\chi_{jj}(\tau_1, \tau_2) = \langle \mathbf{j}(\tau_1), \mathbf{j}(\tau_2) \rangle$, by the Kubo formula. In Fig. 1 we present the experimental optical conductivity function $\sigma_1(\omega)$ of an optimally doped $\text{Bi}_2\text{Sr}_2\text{Ca}_{0.92}\text{Y}_{0.08}\text{Cu}_2\text{O}_{8+\delta}$ single crystal

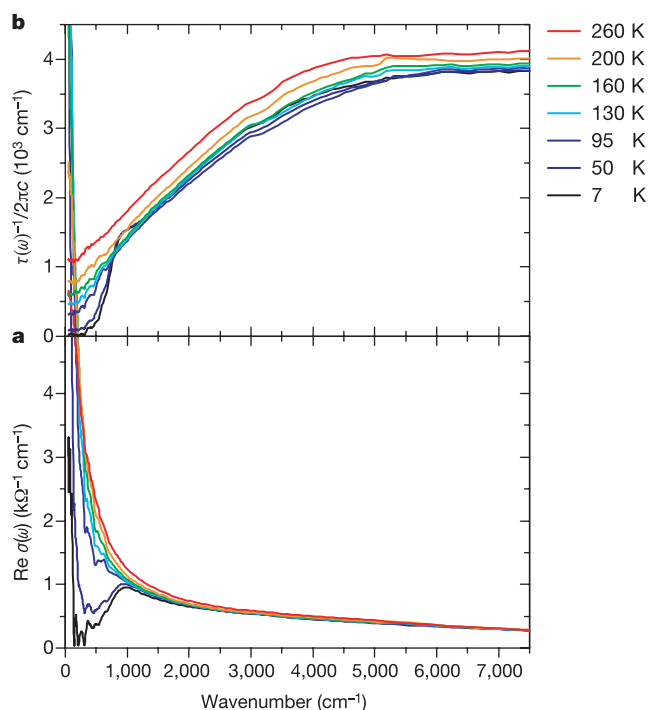


Figure 1 Optical properties along the copper-oxygen planes of $\text{Bi}_2\text{Sr}_2\text{Ca}_{0.92}\text{Y}_{0.08}\text{Cu}_2\text{O}_{8+\delta}$ for a selected number of temperatures. **a**, Optical conductivity and **b**, the frequency dependent scattering rate defined as $1/\tau(\omega) = \text{Re}\{\omega_p^2/4\pi\sigma(\omega)\}$ (see Methods). The relatively high transition temperature ($T_c = 96$ K) of this crystal compared to previous reports on Bi-2212 is caused by the partial substitution of yttrium on the calcium sites.

($T_c = 96$ K; H.E. *et al.*, manuscript in preparation). In order to facilitate comparison with earlier publications^{8–10} we also present $1/\tau(\omega)$ for a number of temperatures, adopting $\omega_p/2\pi c = 19,364\text{ cm}^{-1}$ for the plasma frequency (where c is the velocity of light). The scattering rate $1/\tau(\omega)$ increases approximately linearly as a function of frequency, and when the temperature T is increased, the $1/\tau(\omega)$ curves are shifted vertically proportional to T . The notion that $1/\tau(\omega, T) \sim \omega + T$ in the copper oxides forms one of the centre pieces of the marginal Fermi liquid model^{1,11}, and it has been shown to be approximately correct in a large number of experimental papers^{8–10}. This phenomenology stresses the importance of temperature as the (only) relevant energy scale near optimal doping, which has motivated the idea that optimally doped copper oxides are close to a quantum critical point¹. As can be seen in Fig. 1, $1/\tau(\omega)$ has a negative curvature in the entire infrared region for all temperatures, and it saturates at around $5,000\text{ cm}^{-1}$. Although this departure from linearity may seem to be a minor detail, we will see that it is a direct consequence of the quantum critical scaling of the optical conductivity.

If a quantum phase transition indeed occurs at optimal doping $x = x_c$, then three major frequency regimes of qualitatively different behaviour are expected²: (1) $\omega < T$; (2) $T < \omega < \Omega$; (3) $\Omega < \omega$. As we now report, we find direct indications of these regimes in our optical conductivity data.

Region 1 ($\omega < T$) corresponds to measurement times long compared to the compactification radius of the imaginary time, $L_T = \hbar/k_B T$ (see Methods). Some ramifications have already been discussed above. In addition, Sachdev² showed that in this regime the system exhibits a classical relaxational dynamics characterized by a relaxation time $\tau_r = A L_T$ (A is a numerical prefactor of order 1), reflecting that temperature is the only scale in the system. For the low frequency regime we expect a Drude form $\sigma_1(\omega) = (4\pi)^{-1} \omega_p^2 \tau_r / (1 + \omega^2 \tau_r^2)$, where ω_{pr} is the plasma frequency. Then $T\sigma_1(\omega, T)$ becomes a universal function of ω/T , at least up to a number of order one:

$$\frac{\hbar}{k_B T \sigma_1(\omega, T)} = \frac{4\pi}{A \omega_{pr}^2} \left(1 + A^2 \left(\frac{\hbar \omega}{k_B T} \right)^2 \right) \quad (1)$$

In the inset of Fig. 2 we display $\hbar/(k_B T \sigma_1)$ as a function of $u =$

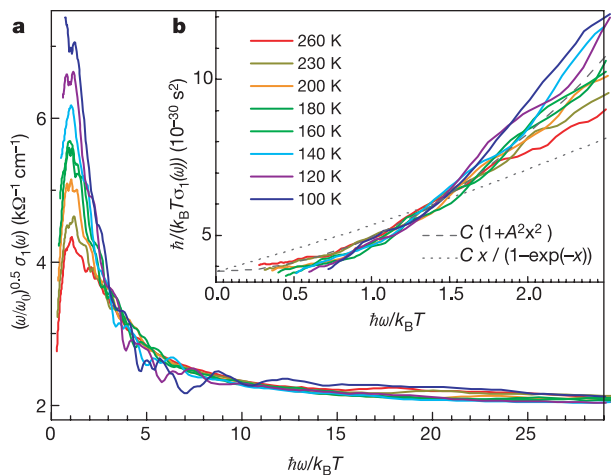


Figure 2 Temperature/frequency scaling behaviour of the real part of the optical conductivity of $\text{Bi}_2\text{Sr}_2\text{Ca}_{0.92}\text{Y}_{0.08}\text{Cu}_2\text{O}_{8+\delta}$. The sample is the same as in Fig. 1. In **a**, the data are plotted as $(\omega/\omega_0)^{0.5} \sigma_1(\omega, T)$. The collapse of all curves on a single curve for $\hbar\omega/k_B T > 3$ demonstrates that in this ω/T -region the conductivity obeys $\sigma_1(\omega, T) = \omega^{-0.5} g(\omega/T) = T^{-0.5} h(\omega/T)$. Note that $g(u) = u^{0.5} h(u)$. In **b**, the data are presented as $\hbar/(k_B T \sigma_1(\omega, T))$, demonstrating that for $\hbar\omega/k_B T < 1.5$ the conductivity obeys $\sigma_1(\omega, T) = T^{-1} f(\omega/T)$.

$\hbar\omega/k_B T$. Clearly the data follow the expected universal behaviour for $u < 1.5$, with $A = 0.77$. The experimental data are in this regard astonishingly consistent with Sachdev's predictions, including $A \approx 1$. From the fitted prefactor we obtain $\omega_{pr}/2\pi c = 9,597\text{ cm}^{-1}$. Above we have already determined the total spectral weight of the free carrier response, $(\omega_p/2\pi c)^2 = 19,364^2\text{ cm}^{-2}$. Hence the classical relaxational response contributes 25% of the free carrier spectral weight. These numbers agree with the results and analysis of Quijada *et al.*⁸. This spectral weight collapses into the condensate peak at $\omega = 0$ when the material becomes superconducting⁸. In Fig. 2 we also display the scaling function proposed by Prelovsek¹², $\sigma_1(\omega) = C(1 - \exp(-\hbar\omega/k_B T))/\omega$. The linear frequency dependence of this formula for $\hbar\omega/k_B T \ll 1$ is clearly absent from the experimental data. The universal dependence of $T\sigma_1(\omega, T)$ on ω/T also contradicts the "cold spot model"¹³, where $T\sigma_1(\omega, T)$ has a universal dependence on ω/T^2 .

In region 2 ($T < \omega < \Omega$) we can probe directly the scale invariance of the quantum critical state. Let us now introduce the scaling relation along the time axis, as follows from elementary consideration. The euclidean (that is, imaginary time) correlator has to be known in minute detail in order to enable the analytical continuation to real (experimental) time. However, in the critical state invariance under scale transformations fixes the functional form of the correlation function completely: It has to be an algebraic function of imaginary time. Hence, it is also an algebraic function of Matsubara frequency $\omega_n = 2\pi n/L_T$, and the analytical continuation is unproblematic: (1) Scale invariance implies that $\sigma_1(\omega)$ and $\sigma_2(\omega)$ have to be algebraic functions of ω , (2) causality forces the exponent to be the same for $\sigma_1(\omega)$ and $\sigma_2(\omega)$, and (3) time reversal

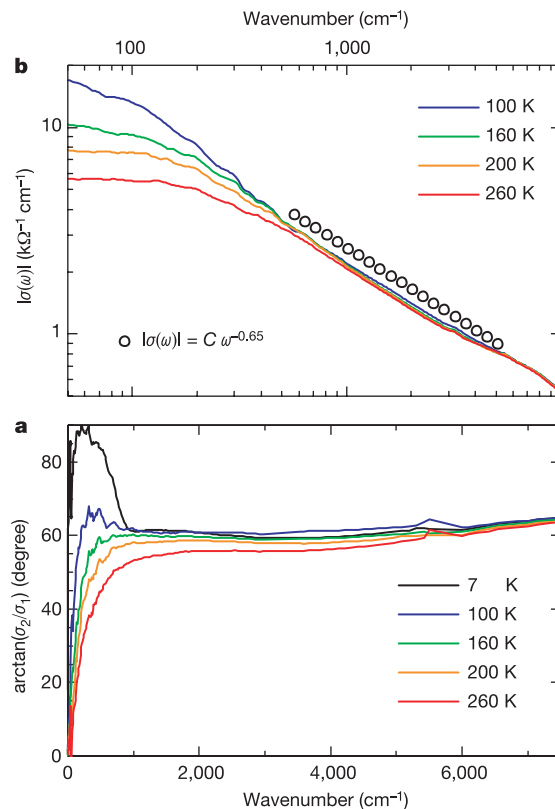


Figure 3 Universal power law of the optical conductivity and the phase angle spectra of optimally doped $\text{Bi}_2\text{Sr}_2\text{Ca}_{0.92}\text{Y}_{0.08}\text{Cu}_2\text{O}_{8+\delta}$. The sample is the same as in Fig. 1. In **a**, the phase function of the optical conductivity, $\text{Arg}(\sigma(\omega))$ is presented. In **b**, the absolute value of the optical conductivity is plotted on a double logarithmic scale. The open symbols correspond to the power law $|\sigma(\omega)| = C\omega^{-0.65}$.

symmetry, implying $\sigma(\omega) = \sigma^*(-\omega)$, fixes the absolute phase of $\sigma(\omega)$. Taken together

$$\sigma(\omega) = C(-i\omega)^{\gamma-2} = C\omega^{\gamma-2}e^{i\pi(1-\gamma/2)} \quad (2)$$

Hence the phase angle relating the reactive and absorptive parts of the conductivity, $\arctan(\sigma_2/\sigma_1) = (2 - \gamma) \cdot 90^\circ$, is frequency independent and should be set by the critical exponent γ . Power-law behaviour of the optical conductivity of the copper oxides has been reported previously^{14,15}, but to our knowledge the relation between the phase and the exponent has not been addressed in the literature. In Fig. 3 we display the frequency dependence of $|\sigma|$ in a log-log plot, and the phase in a linear plot. Although the temperature dependence 'leaks out' to surprisingly high frequency in the latter, the data are remarkably consistent with equation (2) for ω between $k_B T$ and $7,500 \text{ cm}^{-1}$. The observed power law of the conductivity, $|\sigma| = C/\omega^{0.65}$ corresponds to $\gamma = 1.35$, and the value of the phase, $\arctan(\sigma_2/\sigma_1) = 60^\circ \pm 2^\circ$, implies that $\gamma = 1.33 \pm 0.04$. The good consistency of γ obtained from two experimental quantities (that is, the exponent of a power law and the phase) is a strong test of the validity of equation (2). Frequency independence of the phase in region 2 and agreement between the two power laws (one from $\sigma(\omega)$ and the other from the phase angle spectrum) are unique properties of slightly overdoped samples, as demonstrated by Fig. 4, where we present the phase function for optimally doped¹⁶ ($T_c = 88 \text{ K}$), underdoped¹⁶ ($T_c = 66 \text{ K}$), and overdoped ($T_c = 77 \text{ K}$) single crystals of $\text{Bi}_{2.23}\text{Sr}_{1.9}\text{Ca}_{0.96}\text{Cu}_2\text{O}_{8+\delta}$ with different oxygen concentrations¹⁷. The observed trend for different dopings suggests that optimal doping, with $T_c = 88 \text{ K}$, and overdoping, with $T_c = 77 \text{ K}$, are lower and upper carrier concentrations where the optical conductivity obeys equation (2).

Because the phase is constant, the frequency dependent scattering rate $1/\tau(\omega) = \text{Re}\{\omega_p^2/4\pi\sigma(\omega)\} = C\omega^{2-\gamma} = C\omega^{0.65}$ can not be a linear function of frequency (but note that $1/\tau^*(\omega)$ is linear¹⁸, see Methods). Our findings disqualify directly theories that do not incorporate a manifest temporal scale invariance. Luttinger liquids are quantum critical states of matter, and Anderson's results¹⁸ based on one-dimensional physics are therefore of the correct form, equation (2). The exponent $\gamma = 4/3$ is within the range considered by Anderson, but differs significantly from the prediction based on the "cold spot" model¹³, providing $\sigma(\omega) \sim (-i\omega)^{-0.5}$, which corresponds to $\gamma = 3/2$.

Let us now turn to the temperature dependence of the optical conductivity. From Fig. 2 we see that in this region the conductivity

crosses over to a different dependence on ω and T : for $\hbar\omega/k_B T > 3$ the experimental data are seen to collapse onto a curve of the form $\sigma_1(\omega, T) = T^{-0.5} h(\omega/T)$, where $h(u)$ is a universal function. For $u > 3$, $h(u)$ has a weak power-law dependence corresponding to a frequency dependence $\sigma_1(\omega, T) \sim \omega^{-0.65}$. According to the simplest scaling hypothesis, $\sigma_1(\omega, T) \sim T^{-\mu} h'(\omega/T)$ with $h'(u) \rightarrow \text{constant}$ and $h'(u) \rightarrow u^{-\mu}$ in the limits $u \rightarrow 0$ and $u \gg 1$, respectively. Although this energy-temperature scaling is roughly satisfied in the high frequency regime, it is strongly violated at low frequencies, because the success of equation (1) in the regime for $\hbar\omega/k_B T < 1.5$ (see Fig. 2) implies an exponent $\mu = 1$ at low frequencies instead. Bernhoeft has noticed a similar problem in the context of the heavy-fermion critical points¹⁹.

Region 3 ($\omega > \Omega$) necessarily has a different behaviour of the optical conductivity, based on the following simple argument: the spectral weight of the optical conductivity integrated over all frequencies is set by the f-sum rule. However, since $\gamma > 1$, the integration over all frequencies of $\text{Re}\sigma(\omega)$ of the form of equation (2) diverges. Hence we expect a crossover from the constant phase angle $\text{Arg}\sigma(\omega) = (2 - \gamma) \cdot 90^\circ$ to the asymptotic value $\text{Arg}\sigma(\infty) = 90^\circ$. The details of the frequency dependence of $\sigma(\omega)$ at the crossover point Ω depend on the microscopic details of the system. A (non-universal) example of an ultraviolet regularization with the required properties is²⁰ $\sigma(\omega) = (ne^2/m)(-i\omega)^{\gamma-2} \times (\Omega - i\omega)^{1-\gamma}$. Indeed the phase functions show a gradual upward departure from the plateau value for frequencies exceeding $5,000 \text{ cm}^{-1}$ (Fig. 3). This indicates that the 'ultraviolet' cut-off is on the order of 1 eV.

Do our observations shed light on the enigmatic origin of the quantum criticality? In fact, they point unambiguously at three surprising features. First, the current correlator behaves singularly, and this implies that the electromagnetic currents themselves are the order parameter fields responsible for the criticality. Second, the criticality persists up to surprisingly high energies. Third, we have seen that the optical conductivity curves collapse on $\sigma_1(\omega, T) = T^\mu h(\omega/T)$, where $\mu = 1$ for $\omega/T < 1.5$, while $\mu \sim 0.5$ for $\omega/T > 3$. This disqualifies many theoretical proposals. Much of the intuition regarding quantum criticality is based on the rather well understood quantum phase transitions in systems composed of bosons. A canonical example is the insulator-superconductor transition in two space dimensions²¹ where the optical conductivity is found to precisely obey the energy-temperature scaling hypothesis²²,

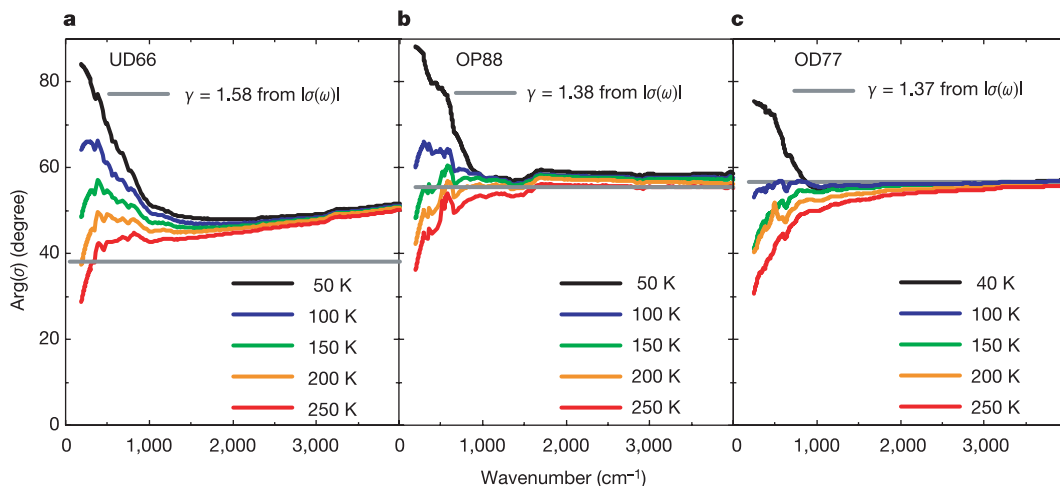


Figure 4 Phase of $\sigma(\omega)$ of $\text{Bi}_{2.23}\text{Sr}_{1.9}\text{Ca}_{0.96}\text{Cu}_2\text{O}_{8+\delta}$ at various doping levels. **a**, Underdoped ($T_c = 66 \text{ K}$); **b**, optimally doped ($T_c = 88 \text{ K}$); and **c**, overdoped ($T_c = 77 \text{ K}$). Solid lines: exponent from $|\sigma(\omega)|$ between $1,000$ and $5,000 \text{ cm}^{-1}$. These crystals were grown at relatively low partial pressures (25 mbar) of oxygen, resulting in

high-quality underdoped crystals with T_c as low as 65 K , and with sharp superconducting transitions $\delta T_c \leq 2.5 \text{ K}$. The optimally doped and overdoped crystals were obtained by post-annealing in oxygen. At optimal doping, T_c turns out to be somewhat lower than the highest values reported in Bi-2212 owing to a slightly different cation ratio.

characterized by a single exponent $\mu = 0$ governing both the frequency and temperature dependences^{2,22,23}. Bosonic theory can be therefore of relevance in electron systems, but it requires the fermionic degrees of freedom to be bound in collective bosonic degrees of freedom at low energy.

In the copper oxides, it appears that the quantum criticality has to do with the restoration of the Fermi-liquid state in the overdoped regime characterized by a large Fermi surface. This implies that fermionic fluctuations play a central role in the quantum critical state, and their role has not yet been clarified theoretically. The absence of a single master curve for all values of ω/T is at variance with notions of quantum critical behaviour, and its understanding may require concepts beyond the standard model of quantum criticality. We close with the speculation that the presence of bosonic fluctuations and fermionic fluctuations in the copper oxides is pivotal in understanding the quantum critical behaviour near optimal doping of the copper oxides. □

Methods

Kubo formula

The Kubo formalism establishes the relation between optical conductivity and current-current correlation function:

$$\sigma(i\omega_n, T) = \frac{Ne^2}{m\omega_n} + \frac{1}{\omega_n} \int_0^{L_T} d\tau e^{i\omega_n \tau} \langle \mathbf{j}(\tau) \mathbf{j}(0) \rangle$$

The first term, corresponding to perfect conductivity, is only relevant in the superconducting state, $\mathbf{j}(\tau)$ is the current operator at (imaginary) time τ , while in the path integral formalism $L_T = \hbar/k_B T$ is the compactification radius of the imaginary time. The angle brackets mean a trace over the thermal distribution at finite temperatures. The integration is over a finite segment of imaginary time, resulting in the optical conductivity at the Matsubara frequencies $i\omega_n = 2\pi i n/L_T$ along the imaginary axis of the complex frequency plane. The conductivity at real frequency ω follows from analytical continuation.

Experimental determination of the optical conductivity

The most direct experimental technique, which provides the optical conductivity and its phase, is spectroscopic ellipsometry. Another popular approach is the measurement of the reflectivity amplitude over a wide frequency region. Kramers–Kronig relations then provide the phase of the reflectivity at each frequency, from which (with the help of Fresnel equations) the real and imaginary part of the dielectric function, $\epsilon(\omega)$, is calculated. We used reflectivity for $50 \text{ cm}^{-1} < \omega/2\pi c < 6,000 \text{ cm}^{-1}$, and ellipsometry for $1,500 \text{ cm}^{-1} < \omega/2\pi c < 36,000 \text{ cm}^{-1}$. This combination allows a very accurate determination of $\epsilon(\omega)$ in the entire frequency range of the reflectivity and ellipsometry spectra. Owing to the off-normal angle of incidence used with ellipsometry, the *ab*-plane pseudo-dielectric function had to be corrected for the *c*-axis admixture. We used previously published²⁴ *c*-axis optical constants of the same compound. The data files were generously supplied to us by S. Tajima. The effect of this correction on the pseudo-dielectric function turns out to be almost negligible, in accordance with Aspnes²⁵.

The optical conductivity, $\sigma(\omega)$, is obtained using the relation $\epsilon(\omega) = \epsilon_\infty + 4\pi i\sigma(\omega)/\omega$, where ϵ_∞ represents the screening by interband transitions. In the copper oxide materials $\epsilon_\infty = 4.5 \pm 0.5$. For $\omega/2\pi c = 5,000 \text{ cm}^{-1}$ an uncertainty of 0.5 of ϵ_∞ propagates to an error of 2° of the phase of $\sigma(\omega)$. This accuracy improves for lower frequencies.

Frequency dependent scattering rate

For an isotropic Fermi liquid, the energy dependent scattering rate of the quasi-particles can be readily obtained from the optical data, using the relation $1/\tau(\omega) = \text{Re}\{\omega_p^2/4\pi\sigma(\omega)\}$. In spite of the fact that the notion of a quasi-particle in the spirit of Landau's Fermi liquid is far from being established for the copper oxides, during the past 15 years it has become a rather common practice to represent infrared data of these materials as $1/\tau(\omega)$. The dynamical mass is defined as $m^*(\omega)/m = \text{Im}\{\omega_p^2/4\pi\sigma(\omega)\}$. To obtain absolute numbers for $1/\tau(\omega)$ and $m^*(\omega)/m$ from the experimental optical conductivity, a value of the plasma frequency, ω_p , must be adopted. With our value of ω_p the dynamical mass converges to 1 for $\omega \rightarrow \infty$. Sometimes the renormalized scattering rate, $\tau^*(\omega)^{-1} = \tau(\omega)^{-1} m/m^*(\omega) = \omega\sigma_1(\omega)/\sigma_2(\omega)$, is reported instead of $\tau(\omega)$. If the frequency dependence of the conductivity is a power law, $\sigma(\omega) = (-i\omega)^{-\gamma}$, then $1/\tau^*(\omega) = -\omega \cotan(\pi\gamma/2)$, which is a linear function of frequency¹⁸. The value of the slope reveals the exponent, and corresponds to the phase of the conductivity displayed in Figs 3 and 4.

Received 10 June; accepted 5 August 2003; doi:10.1038/nature01978.

- Varma, C. M., Nussinov, Z. & van Saarloos, W. Singular or non-Fermi Liquids. *Phys. Rep.* **361**, 267–417 (2002).
- Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, Cambridge, 1999).
- Saxena, S. S. *et al.* Superconductivity on the border of itinerant-electron ferromagnetism in UGe₂. *Nature* **406**, 587–592 (2000).
- Tallon, J. L. & Loram, J. W. The doping dependence of T^* —what is the real high- T_c phase diagram? *Physica C* **349**, 53–68 (2001).
- Orenstein, J. & Millis, A. J. Advances in the physics of high-temperature superconductivity. *Science* **288**, 468–474 (2000).

- Kaminski, A. *et al.* Spontaneous breaking of time-reversal symmetry in the pseudogap state of a high- T_c superconductor. *Nature* **416**, 610–613 (2002).
- Mook, H. A. *et al.* Magnetic order in YBa₂Cu₃O_{6+x} superconductors. *Phys. Rev. B* **66**, 144513 (2002).
- Quijada, M. A. *et al.* Anisotropy in the *ab*-plane optical properties of Bi₂Sr₂CaCu₂O₈ single-domain crystals. *Phys. Rev. B* **60**, 14917–14934 (1999).
- Puchkov, A. V., Basov, D. N. & Timusk, T. The pseudogap state in high- T_c superconductors: An infrared study. *J. Phys.* **8**, 10049–10082 (1996).
- Santander-Syro, A. F. *et al.* Absence of a loss of in-plane infrared spectral weight in the pseudogap regime of Bi₂Sr₂CaCu₂O_{8+x}. *Phys. Rev. Lett.* **88**, 097005–097008 (2002).
- Littlewood, P. B. & Varma, C. M. Phenomenology of the normal and superconducting states of a marginal Fermi liquid. *J. Appl. Phys.* **69**, 4979–4984 (1991).
- Prellovsek, P. On the universal optical conductivity and single-particle relaxation in cuprates. *Europhys. Lett.* **53**, 228–232 (2001).
- Ioffe, L. B. & Millis, A. J. Zone-diagonal-dominated transport in high- T_c cuprates. *Phys. Rev. B* **58**, 11631–11637 (1998).
- Schlesinger, Z. *et al.* Superconducting energy gap and normal-state conductivity of a single-domain YBa₂Cu₃O₇ crystal. *Phys. Rev. Lett.* **65**, 801–804 (1990).
- El Azrak, A. *et al.* Infrared properties of Yba₂Cu₃O₇ and Bi₂Sr₂Ca_{n-1}Cu_nO_{2n+4} thin films. *Phys. Rev. B* **49**, 9846–9856 (1994).
- Molegraaf, H. J. A., Presura, C., van der Marel, D., Kes, P. H. & Li, M. Superconductivity-induced transfer of in-plane spectral weight in Bi₂Sr₂CaCu₂O_{8+x}. *Science* **295**, 2239–2241 (2002).
- Li, M., van der Beek, C. J., Konczykowski, M., Menovsky, A. A. & Kes, P. H. Superconducting properties of strongly underdoped Bi₂Sr₂CaCu₂O_{8+x} single crystals. *Phys. Rev. B* **66**, 024502 (2002).
- Anderson, P. W. Infrared conductivity of cuprate metals: Detailed fit using Luttinger-liquid theory. *Phys. Rev. B* **55**, 11785–11788 (1997).
- Bernhoeft, N. An analysis of the dynamical magnetic susceptibility in non-Fermi liquids. *J. Phys. Condens. Matter* **13**, R771–R816 (2001).
- van der Marel, D. Anisotropy of the optical conductivity of high T_c cuprates. *Phys. Rev. B* **60**, R765–R768 (1999).
- Liu, Y. & Goldman, A. M. Superconductor-insulator transitions in two dimensions. *Mod. Phys. Lett. B* **8**, 277–309 (1994).
- Fisher, M. P. A., Grinstein, G. & Girvin, S. M. Presence of a quantum diffusion in two dimensions: Universal resistance at the superconductor-insulator transition. *Phys. Rev. Lett.* **64**, 587–590 (1990).
- Damle, K. & Sachdev, S. Nonzero-temperature transport near quantum critical points. *Phys. Rev. B* **56**, 8714–8733 (1997).
- Tajima, S., Gu, D. G., Miyamoto, S., Odagawa, A. & Koshizuka, N. Optical evidence for strong anisotropy in the normal and superconducting states in Bi₂Sr₂CaCu₂O_{8+x}. *Phys. Rev. B* **48**, 16164–16167 (1993).
- Aspnes, D. E. Approximate solution of ellipsometric equations for optically biaxial crystals. *J. Opt. Soc. Am.* **70**, 1275–1277 (1980).

Acknowledgements We thank C. M. Varma, P. Prellovsek, C. Pepin, S. Sachdev and A. Tselvik for comments during the preparation of this work, and N. Kaneko for technical assistance. This investigation was supported by the Netherlands Foundation for Fundamental Research on Matter (FOM) with financial aid from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). The crystal growth work at Stanford University was supported by the Department of Energy's Office of Basic Energy Sciences, Division of Materials Science.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.v.d.M. (dirk.vandermarel@physics.unige.ch).

High-performance thin-film transistors using semiconductor nanowires and nanoribbons

Xiangfeng Duan^{1,2}, Chunming Niu^{1,2}, Vijendra Sahi^{1,2}, Jian Chen², J. Wallace Parce², Stephen Empedocles² & Jay L. Goldman^{1,2}

¹Nanosys, Inc., Advanced Research Center, 200 Boston Ave, Medford, Massachusetts 02155, USA

²Nanosys, Inc., 2625 Hanover Street, Palo Alto, California 94304, USA

Thin-film transistors (TFTs) are the fundamental building blocks for the rapidly growing field of macroelectronics^{1,2}. The use of plastic substrates is also increasing in importance owing to their light weight, flexibility, shock resistance and low cost^{3,4}. Current polycrystalline-Si TFT technology is difficult to implement on plastics because of the high process temperatures required^{1,2}. Amorphous-Si and organic semiconductor^{5,6} TFTs, which can be

processed at lower temperatures, but are limited by poor carrier mobility. As a result, applications that require even modest computation, control or communication functions on plastics cannot be addressed by existing TFT technology. Alternative semiconductor materials^{7,8} that could form TFTs with performance comparable to or better than polycrystalline or single-crystal Si, and which can be processed at low temperatures over large-area plastic substrates, should not only improve the existing technologies, but also enable new applications in flexible, wearable and disposable electronics. Here we report the fabrication of TFTs using oriented Si nanowire thin films or CdS nanoribbons as semiconducting channels. We show that high-performance TFTs can be produced on various substrates, including plastics, using a low-temperature assembly process. Our approach is general to a broad range of materials including high-mobility materials (such as InAs or InP).

Individual semiconductor nanowires (NWs)^{9,10} and single-walled carbon nanotubes^{11–13} have been used for nanoscale field-effect transistors (FETs) with performance comparable to or exceeding that of the single-crystal materials. In particular, carrier mobility values of $300 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ have been demonstrated for p-type Si NWs⁹, $2,000\text{--}4,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for n-type InP NWs¹⁰ and up to $20,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ for single-walled carbon nanotubes¹³. These nanofETs promise to push Moore's law to the limit with unprecedented performance¹⁰.

Here we take nanomaterial-enabled electronics in a new direction: we exploit nanomaterials not for the next generation of nanoelectronics, but for high-performance macroelectronics. In short, we assemble NWs into oriented NW thin films to yield a novel electronic substrate; this substrate is processed using standard methods to produce NW-TFTs with conducting channels formed by multiple parallel single-crystal NW paths. In such NW-TFTs, charges travel from source to drain within single crystals, thus ensuring high carrier mobility.

Figure 1a illustrates our NW-TFT fabrication process. We synthesized p-type Si-NWs with controlled diameters using a previously

reported approach¹⁴. NWs have a core-shell structure, with a single-crystal core surrounded by an amorphous silicon oxide shell of 1–3 nm thickness (Supplementary Fig. S1). The NWs were then dispersed into solution, and assembled onto the surface of the chosen substrate using a flow-directed alignment method¹⁵ to produce an oriented NW thin film. An optical micrograph (Fig. 1b) of the NW thin film shows that the film consists of a monolayer of NWs oriented in parallel with an average NW spacing of 500–1,000 nm. The NW spacing is controlled by varying the NW solution concentration and flow time. Other approaches (for example, a Langmuir–Blodgett film) may also be used to obtain nearly close-packed NW thin films^{16,17}. Oriented-NW deposition can readily be achieved over a 4-inch wafer and potentially at larger scales (Supplementary Fig. S2). The NW thin film was then processed using standard lithography followed by metallization to define source and drain electrodes and yield TFTs (Fig. 1c, d). For initial study, the TFTs had a simple back-gated device configuration on a silicon substrate (Fig. 1c inset), where underlying silicon was used as the back gate, 100-nm-thick silicon nitride (SiN_x) as the gate dielectric, and Ti/Au film as the source and drain electrodes.

Drain current (I_{DS}) versus drain–source voltage (V_{DS}) relations (Fig. 2a) at various gate voltages (V_{GS}) for a NW-TFT show typical accumulation mode p-channel transistor behaviour¹⁸, as I_{DS} increases linearly with V_{DS} at low V_{DS} , and saturates at higher V_{DS} . Upon application of negative V_{GS} , I_{DS} increases as the majority carrier (hole) density increases in the channel. Applying a positive V_{GS} depletes holes in the channel and turns the device off. The plot of $-I_{\text{DS}}$ versus V_{GS} (Fig. 2b) at a constant $V_{\text{DS}} = -1 \text{ V}$ shows little current when the V_{GS} is more positive than a threshold voltage (V_{th}), and I_{DS} increases nearly linearly when the V_{GS} increases in the negative direction. Extrapolation of the linear region results in a V_{th} of 0.45 V.

For macroelectronic applications, a number of key transistor parameters, including transconductance, mobility, on/off ratio, threshold voltage, and subthreshold swing, dictate TFT performance¹⁸. High transconductance is a critical measure of transistor

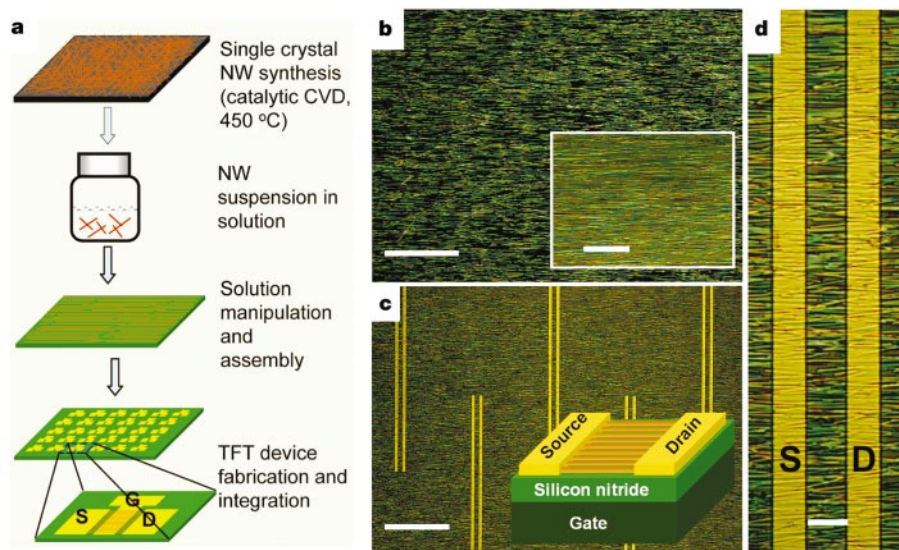


Figure 1 Nanowire TFT fabrication. **a**, Diagrams illustrating the NW-TFT fabrication process. In our approach, high-quality single-crystal NW materials are synthesized at high temperature using a catalytic chemical vapour deposition processes, and are then harvested and dispersed in a solution and aligned on the desired substrate to form a densely-packed oriented NW thin film that is further processed via standard methods to form TFTs with the conducting channel parallel to the wire axis. **b**, Optical micrograph of flow aligned NW thin film. Scale bar, 80 μm . Inset, a picture at higher magnification. Scale bar, 20 μm . The micrographs show that the NW thin film is nearly a monolayer of NWs, but

occasionally a few NWs cross over each other. The average space between parallel NW arrays is estimated to be $\sim 540 \text{ nm}$. **c**, Optical micrograph of NW-TFTs fabricated on NW thin films. The whole substrate is covered with a NW thin film, and each pair of metal (Ti/Au) electrodes (yellow) corresponds to source–drain electrodes for a NW-TFT. Scale bar, 100 μm . Inset, a schematic view of the NW-TFT configuration. **d**, Optical micrograph of a NW-TFT, where parallel arrays of NW are clearly seen to bridge the full distance from the source to the drain electrodes. Scale bar, 5 μm .

performance, and determines voltage gains of transistor-based devices including amplifiers and logic circuits. The slope in the linear region of $-I_{DS}$ versus V_{GS} gives a transconductance $g_m = dI_{DS}/dV_{GS} \approx 11 \mu S$ at $V_{DS} = -1$ V. Assuming the effective channel width equals the NW diameter (d) multiplied by the number (N) of

NWs, $W_{eff} = Nd = 1.8 \mu m$, we obtain a normalized transconductance of $\sim 6 \mu S \mu m^{-1}$; this is significantly better than that of organic⁶, amorphous (a)-Si ($< 0.01 \mu S \mu m^{-1}$)¹, and p-channel polycrystalline (poly)-Si TFTs (~ 0.2 – $0.8 \mu S \mu m^{-1}$)¹⁹, and is comparable to that of single-crystal p-channel silicon-on-insulator (SOI) MOSFETs (~ 5 – $12 \mu S \mu m^{-1}$)²⁰. A normalized transconductance of $\sim 0.09 \mu S \mu m^{-1}$ is calculated if the physical channel width of the device is used, which is still significantly better than a-Si TFTs, and nearly comparable to poly-Si TFTs. The value can be further increased with improved assembly approach^{16,17}. For a fair comparison, all the transconductance values have been scaled to devices with the same dimension according to a scaling law¹⁸. We note that the transconductance can further be improved using thinner dielectrics of higher dielectric constant¹².

To estimate the carrier mobility, the device was modelled using standard MOSFET equations. In the low-bias linear region of the I_{DS} – V_{DS} curves, the hole mobility μ_h can be deduced from $G_{DS} = I_{DS}/V_{DS} = \mu_h C_G (V_{GS} - V_{th} - V_{DS}/2)/L^2$, where C_G is the gate capacitance and L is the channel length. Using a calculated C_G (see Methods), we deduced $\mu_h \approx 119 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, which is significantly larger than those of organic, a-Si TFTs ($< 1 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) and comparable to the best reported values for p-type poly-Si TFTs ($\sim 120 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$)² and that of p-type SOI MOSFETs ($\sim 180 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$)²⁰. Additionally, it is possible to further improve the carrier mobility; for example, by decreasing the doping level and/or minimizing the trapping states on the NW surface^{1,9}.

The curve of $-I_{DS}$ versus V_{GS} on an exponential scale (Fig. 2b inset) shows that I_{DS} decreases exponentially below V_{th} and that the transistor has an on-off current ratio of nearly 10^8 . This is, to our knowledge, the largest on-off ratio reported for transistors assembled from chemically synthesized nanomaterials, and is comparable to that of single-crystal silicon devices¹⁸. The subthreshold swing $S = -dV_{GS}/d\ln|I_{DS}|$ is ~ 600 mV per decade for this device. In conventional MOSFETs, subthreshold swing depends on the ratio of the gate capacitance to other capacitances such as interface trap state capacitance, and has a theoretical limit of ~ 60 mV per decade at room temperature¹⁸. A small subthreshold swing is usually desired for low-threshold-voltage, low-power operation. The subthreshold swing of ~ 600 mV per decade in our device is significantly better than those of organic⁶ or a-Si TFTs¹ that typically range from one to many volts per decade, but is substantially larger than the best values in poly-Si TFTs (~ 200 mV per decade)¹ and single-crystal silicon devices (~ 70 mV per decade)¹⁸. We believe the relatively large subthreshold swing in our device is mainly due to the existence of surface trapping states¹ and a geometric effect (Supplementary Fig. S3b), which can be improved dramatically by passivating the surface (for example, hydrogenation^{1,21} or using a core-shell structure²²) and/or using a top- or surrounding-gated structure with high- k dielectrics¹². Importantly, a subthreshold swing as small as ~ 70 mV per decade has been demonstrated using a surrounding conformal electrolyte gate¹³ (X.D., C.N. and V.S., unpublished results).

To study the reproducibility and practical viability of this technology, we have performed tests on 20 NW-TFTs. The threshold voltage distribution (Fig. 2c) and gaussian fitting show a standard deviation of only 0.22 V. Considering that the device configuration and fabrication process have not been optimized, such a tight distribution is very encouraging. We note that major device parameters are independent of the number of NWs in the conducting channel except the drain-source on-current, which scales linearly with the effective channel width (Fig. 2d). Reproducible and predictable assembly of NW-TFTs with designed device parameters will be critical for actual circuit design in future applications.

Because the NW synthesis is independent of the device substrate, our process represents a general approach for exploiting a broad range of materials (including III-V and II-VI group semiconduc-

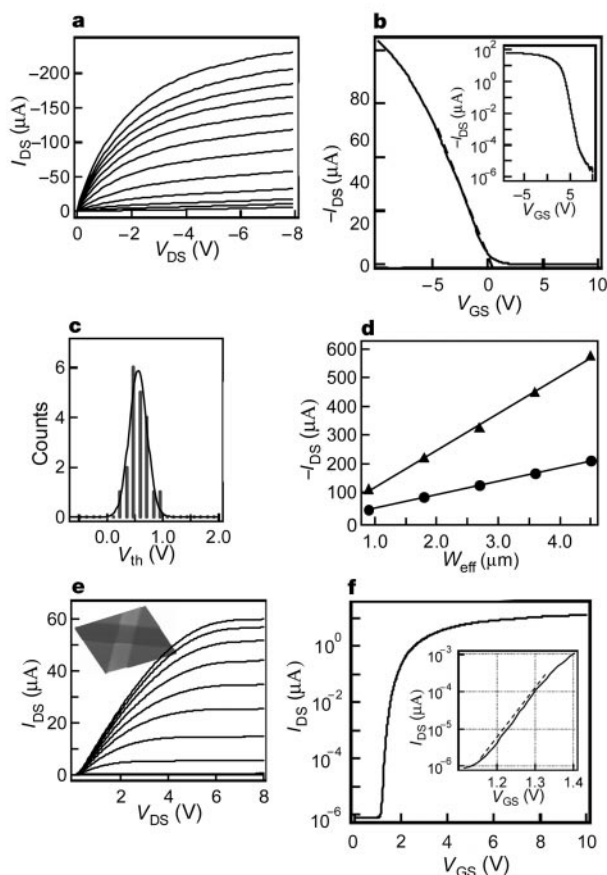


Figure 2 p-channel Si NW-TFT and n-channel CdS nanoribbon TFT. **a**, Drain current (I_{DS}) versus drain-source voltage (V_{DS}) (I_{DS} – V_{DS}) at increasing gate voltages (V_{GS}) in steps of 1 V starting from the top at $V_{GS} = -10$ V. The TFT consists of 91 20-nm-diameter NWs in parallel with a 5- μm channel length. **b**, $-I_{DS}$ versus V_{GS} at $V_{DS} = -1$ V. Linear extrapolation results in a threshold voltage (V_{th}) of ~ 0.45 V. Inset, $-I_{DS}$ versus V_{GS} at $V_{DS} = -1$ V on an exponential scale, highlighting the on-off ratio of nearly 10^8 , and the subthreshold swing of about 600 mV per decade. The linear plot data (main panel) were collected at a V_{GS} sweep rate of 500 mV s^{-1} , and the exponential plot data (inset) were collected at a V_{GS} sweep rate of 15 mV s^{-1} to minimize capacitive charging currents at the higher gate voltages. The apparent threshold voltage in the inset is shifted to 3.5 V due to a hysteresis effect (Supplementary Fig. S5). **c**, Histogram of threshold voltage (V_{th}) distribution from 20 NW-TFT devices shows high device-to-device reproducibility and a tight distribution. Gaussian fitting (curve) shows a standard deviation of only 0.22 V. The threshold voltage value was derived from linear extrapolation of the I_{DS} – V_{GS} relation at $V_{DS} = -1$ V. The I_{DS} – V_{GS} relation for all the devices was taken under exactly the same conditions, with a V_{GS} sweep rate of 500 mV s^{-1} . **d**, Linear-scale relation for the drain current when the device is turned on ($V_{GS} = -10$ V). The circles and triangles represent the on-state current as a function of effective channel width measured at $V_{DS} = -1$ V and -8 V, respectively. The effective channel width corresponds to the product of the average diameters of the NWs and the number of the NWs in the channel. **e**, I_{DS} – V_{DS} at variable gate voltages (V_{GS}) starting from top at $V_{GS} = +10$ V decreasing in steps of 1 V. The TFT consists of a nanoribbon of 0.7- μm width and 4.3- μm channel length. Inset, a 3D atomic force microscope topographic image of a nanoribbon TFT. **f**, I_{DS} – V_{GS} relation at $V_{DS} = 1$ V on an exponential scale shows an on-off ratio greater than 10^7 . Inset, a zoomed-in plot in the subthreshold region that highlights a subthreshold swing as small as 70 mV per decade.

tors²³) as the TFT channel materials, and thus opens new opportunities not possible with current poly-Si technologies. As an example, we now show that high-performance TFTs can also be readily assembled from CdS nanoribbons. CdS is an excellent material for optical as well as electronic applications owing to its intrinsically low surface trapping states²⁴, and is one of the earliest materials used for TFTs²⁵. Single-crystal CdS nanoribbons (Supplementary Fig. S4) were synthesized using a vacuum-vapour-transport method. These nanoribbons are particularly interesting candidates for assembling TFTs because their unique physical morphology closely resembles that of conventional single-crystal thin film. CdS nanoribbon TFTs with a single-crystal conducting channel (Fig. 2e inset) were fabricated with an approach similar to that described above.

Electrical transport measurements show typical n-channel transistor characteristics, consistent with previous studies on CdS bulk materials and NWs²⁴. The I_{DS} – V_{DS} relation at different gate voltages (Fig. 2e) shows a linear region at low V_{DS} and saturates at a higher V_{DS} . The I_{DS} – V_{GS} relation at a V_{DS} of 1 V shows nearly linear behaviour above a threshold V_{GS} of 2.0 V. The slope in the linear region gives a transconductance of $\sim 2.4 \mu\text{S}$ at $V_{DS} = 1$ V. With the calculated capacitance (see Methods), we deduce electron mobility

of $\sim 283 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ using $I_{DS}/V_{DS} = \mu_e C_G (V_{GS} - V_{th} - V_{DS}/2)/L^2$. Importantly, this mobility value matches closely with that of single-crystal CdS material ($\sim 300\text{--}350 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$)²⁶. Furthermore, the exponential plot of I_{DS} – V_{GS} gives an on-off ratio $>10^7$ and a subthreshold swing as small as 70 mV per decade (Fig. 2f and inset), approaching the theoretical limit of 60 mV per decade. The high carrier mobility and small subthreshold swing can be largely attributed to excellent crystalline quality (Supplementary Fig. S4), low-surface states in CdS nanoribbons and the absence of geometrical effects like that in Si NW-TFTs.

A unique feature of the NW-TFT concept is that the entire NW-TFT fabrication process was performed essentially at room temperature, except for the NW synthesis step which is separate from the device fabrication. Therefore, the assembly of high-performance NW-TFTs could be readily applied to low-cost glass and plastic substrates. To demonstrate NW-TFTs on plastics, we have adopted a different device configuration (Fig. 3a). In the plastic device, a locally patterned Cr/Au electrode was used as the gate and electron-beam (e-beam)-evaporated aluminium oxide (AlO_x) was used as the gate dielectric (Fig. 3b).

I_{DS} – V_{DS} curves of a NW-TFT on plastic (Fig. 3c) show a similar behaviour to that of devices on SiN_x/Si substrate. The I_{DS} – V_{GS} relation (Fig. 3d) shows a threshold voltage of ~ 3.0 V, an on-off ratio $>10^5$, and a subthreshold swing of 500–800 mV per decade, which to our knowledge are among the best characteristics reported for TFTs on plastic. The relatively smaller on-off ratio is due to (1) lower on-current resulting from un-optimized local-gate device configuration, and (2) higher off-current limited by gate leakage current caused by the low quality of e-beam-evaporated AlO_x dielectrics; the on-off ratio can be significantly increased with improved device configuration and advanced core-shell NW structure (Supplementary Fig. S3). A major motivation driving the plastic electronic research is mechanical flexibility. Importantly, slight flexing (radius of curvature ~ 55 mm) of the plastic with a NW-TFT device does not significantly change the device behaviour (black versus red curve in Fig. 3d and inset). The linear region in the I_{DS} – V_{GS} relation gives a transconductance of $0.45 \mu\text{S}$ at $V_{DS} = -1$ V (Fig. 3d), but it is hard to extract the hole mobility in the device owing to the difficulties in estimating the gate capacitance of the local-gated device configuration. From electrical measurement of a NW-TFT on plastic with an electrolyte gate¹³ (X.D., C.N. and V.S., unpublished results), we deduced the hole mobility $\mu_h \approx 123 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, which is in good agreement with the mobility obtained for similar devices on SiN_x/Si substrate with similar NWs.

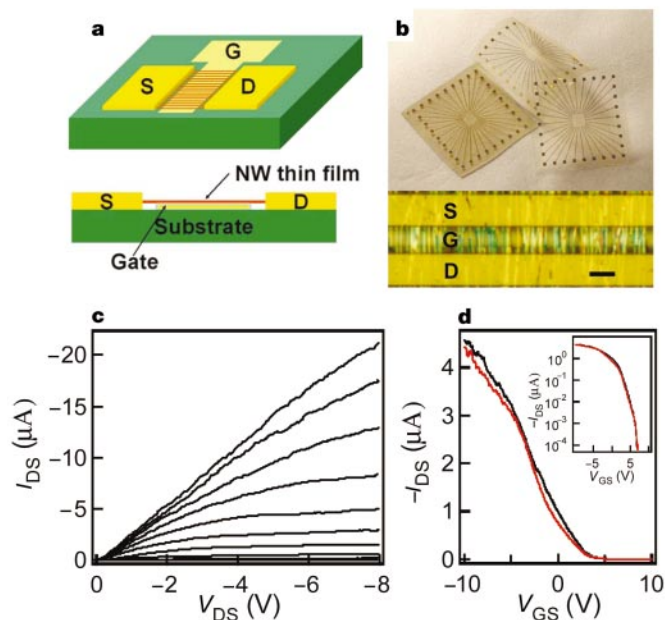


Figure 3 NW-TFTs on plastic. **a**, Diagrams show the device configuration of a locally back-gated NW-TFT on plastic: top, top view; bottom, cross-sectional view. To fabricate the device, a layer of 1–2 μm -thick SU-8 (MicroChem Corp.) photo-resist was first spin cast and cured on a polyetheretherketone (PEEK) sheet (125 μm thick, Goodfellow Inc.) to ensure a microscopically smooth surface. Cr/Au (10/30 nm) strips were then defined as the gate arrays, and a 30-nm layer of aluminium oxide was deposited as gate dielectric using e-beam evaporation. Lastly, the aligned NW thin film was deposited onto the surface, and Ti/Au (60/80 nm) source–drain electrodes were defined to form the TFTs. **b**, Top, a picture of plastic (PEEK) devices with NW-TFTs. The plastic devices show high mechanical flexibility. Bottom, optical micrograph of a locally gated NW-TFT. Scale bar, 5 μm . **c**, I_{DS} – V_{DS} relation at variable V_{GS} starting from the top at $V_{GS} = -8$ V and increasing in steps of 1 V. The TFT consists of 17 NWs of 40-nm diameter in parallel with a 6- μm channel length and 3- μm gate length. **d**, $-I_{DS}$ – V_{GS} relation at $V_{DS} = -1$ V. Inset, $-I_{DS}$ – V_{GS} relation at $V_{DS} = -1$ V on an exponential scale, which highlights the on-off ratio of more than 10^5 and a subthreshold swing of 500–800 mV per decade. The black and red curves show the transfer characteristics of the same device before and after slight flexing of the plastic substrate (radius of curvature ~ 55 mm), demonstrating the mechanical flexibility of NW TFTs on plastics.

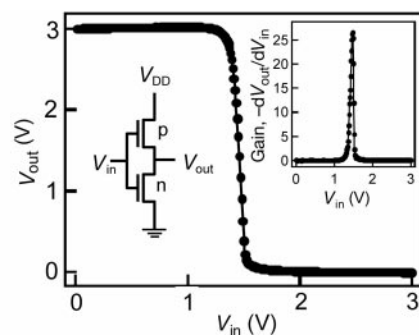


Figure 4 Complementary inverter constructed from NW and nanoribbon TFTs. The output–input (V_{out} versus V_{in}) curve shows a sharp inversion of the output signal. Left inset, the circuit of the complementary inverter formed by connecting a p-channel Si NW-TFT (consisting of 15 NWs in parallel) and an n-channel CdS nanoribbon TFT in series. Right inset, the gain (absolute value of the derivative of V_{out} – V_{in} relation) of the inverter, highlighting a large voltage gain of 27.

The dissociation of semiconductor material growth and device fabrication in our approach allows us to flexibly combine different high-performance semiconductor materials together on a single substrate to achieve (much) improved function and performance, in a way not possible with other technologies. For example, a complementary inverter (a logic NOT gate) was constructed by connecting a p-channel Si NW TFT and an n-channel CdS nanoribbon TFT in series (Fig. 4 left inset)²⁷. The output–input ($V_{\text{out}}-V_{\text{in}}$) voltage response (Fig. 4) of the inverter shows constant high voltage output with low input. When the input is increased to about 1.5 V, the output quickly turns to 0 V and maintains a low state at higher input voltages. Most significantly, the complementary inverter exhibits a high voltage gain. Differentiation of the measured $V_{\text{out}}-V_{\text{in}}$ relation reveals a voltage gain as large as 27 (Fig. 4 right inset). Such a large gain demonstrates high performance of our devices, and is critical for interconnection of arrays of logic circuits for macroelectronic applications without the need for signal restoration at each stage.

We have demonstrated an approach to macroelectronics based on solution-assembled NW or nanoribbon thin films. This approach offers a potential technology platform for high-performance TFTs from high-mobility materials (for example, InP or InAs)^{10,23}, and for novel large-area optoelectronics^{24,28} from II–VI and III–V group optically active NW²³ materials on flexible substrates. Significantly, low-cost, low-temperature processes using microcontact²⁹ or ink-jet printing technology³⁰ may be employed to produce high-performance flexible macroelectronics. We believe that the present approach will affect existing applications and enable new opportunities in flexible, wearable and disposable electronics. □

Methods

Material syntheses

The p-type Si NWs with controlled diameter of 20 or 40 nm were synthesized by thermal deposition of SiH_4 and B_2H_6 using commercially available mono-dispersed gold colloid particles (British Biocell International Ltd) as the catalysts¹⁴. The growth was typically carried out in a computer-controlled pilot production scale reactor in a 8-inch tube furnace at a temperature between 420 and 480 °C, a total pressure of 30 torr, and a silane partial pressure of approximately 2 torr, for a period of 40 min. The SiH_4 to B_2H_6 ratio can be varied to control the doping level, and a ratio of ~6,400:1 was used in synthesizing NWs for this study. The doping concentration of the NW was estimated to be $\sim 4 \times 10^{17} \text{ cm}^{-3}$ from electrical characterizations. This relatively high doping concentration was used to ensure a good contact between NWs and source and drain electrodes without a high-temperature annealing process.

CdS nanoribbons were synthesized using a vacuum vapour transport approach. Specifically, a small amount of CdS powder (~100 mg) was transferred into one end of a vacuum tube and sealed. The vacuum tube was heated such that the end with CdS powder was maintained at 900 °C, while the other end was kept at a temperature ~50 °C lower. Within two hours, most of the CdS was transported to the cooler end and deposited on the tube wall. The resulting materials are predominantly nanoribbons having thicknesses of 30–150 nm, widths of 0.5–5 μm , and lengths of 10–200 μm (Supplementary Fig. S4).

Device fabrication and characterization

The as-made NWs or nanoribbons were dispersed into ethanol solution by ultrasonication, and assembled on to chosen substrate for device fabrication using a fluidic flow alignment approach¹⁵. For a global-back-gated device on SiN_x/Si substrate, NWs or nanoribbons were first aligned to the substrate to form a thin film, which is further subjected to photolithography or e-beam lithography to define the source and drain electrode. Source and drain electrodes were metallized with Ti/Au (60/80 nm) film deposited using an e-beam evaporator, without further thermal annealing. Prior to metallization, the native oxide was etched using a dilute (6%) buffered hydrofluoric acid solution. For local-gated devices on plastics, a layer of 1–2- μm -thick SU-8 (MicroChem Corp.) photo-resist was first spin cast and cured on a PEEK sheet (125 μm thick, Goodfellow Inc.) to ensure a microscopically smooth surface. Cr/Au (10/30 nm) strips were then defined as the gate arrays, and a 30-nm layer of aluminium oxide was deposited as gate dielectric using e-beam evaporation. Lastly, the aligned NW thin film was deposited onto the surface, and Ti/Au (60/80 nm) source–drain electrodes were defined to form the TFTs.

All electrical characterizations were carried out in air in a dark box at room temperature. For comparison of device performance with other technology (for example, a-Si, poly-Si and SOI structure), all devices were scaled to have the same dimension where applicable. For a global-back-gated Si NW-TFT (Fig. 2 a–d), the gate capacitance includes the capacitance of the SiN_x dielectric on the substrate and that of the

silicon oxide shell. It is, however, nontrivial to calculate these capacitances owing to the complex geometry. The gate capacitance was computed using a 3D finite element package (MetaMesh and HiPhi from Field Precision: <http://www.fieldp.com>). Results are converged within 1%. Numerical simulation yields a total capacitance of ~28 fF for the device shown Fig. 2a and b. For CdS nanoribbon TFTs (Fig. 2e, f), assuming a parallel plate model, we calculated the gate capacitance to be 2.0 fF using $C_G = \epsilon \epsilon_0 L W/h$, where ϵ is dielectric constant, L and W are channel length and width, and h is the dielectric thickness.

Received 26 May; accepted 20 August 2003; doi:10.1038/nature01996.

1. Madelung, O. (ed.) *Technology and Applications of Amorphous Silicon* (Springer, Berlin, 2000).
2. Ucjikoga, S. Low-temperature polycrystalline silicon thin-film transistor technologies for system-on-glass displays. *MRS Bull.* **27**, 881–886 (2002).
3. Rogers, J. A. *et al.* Paper-like electronic displays: Large-area rubber-stamped plastics sheet of electronics and microencapsulated electronic inks. *Proc. Natl Acad. Sci. USA* **98**, 4835–4840 (2001).
4. Lee, M. J., Judge, C. P. & Wright, S. W. Thin film transistors for displays on plastic substrate. *Solid State Electron.* **44**, 1431–1434 (2000).
5. Garnier, F., Hajlaoui, R., Yassar, A. & Srivastava, P. All-polymer field-effect transistor realized by printing techniques. *Science* **265**, 1684–1686 (1994).
6. Dimitrakopoulos, C. D. & Mascaro, D. J. Organic thin-film transistors: A review of recent advances. *IBM J. Res. Dev.* **45**, 11–27 (2001).
7. Ridley, B. A., Nivi, B. & Jacobson, J. M. All-inorganic field effect transistors fabricated by printing. *Science* **286**, 746–749 (1999).
8. Kagan, C. R., Mitzi, D. B. & Dimitrakopoulos, C. D. Organic-inorganic hybrid materials as semiconducting channels in thin-film field-effect transistors. *Science* **286**, 945–947 (1999).
9. Cui, Y., Zhong, Z., Wang, D., Wang, W. & Lieber, C. M. High performance silicon nanowire field effect transistors. *Nano Lett.* **3**, 149–152 (2003).
10. Duan, X., Huang, Y., Cui, Y. & Lieber, C. M. in *Molecular Nanoelectronics* (eds Reed, M. A. & Lee, T.) (American Scientific Publishers, Stevenson Ranch, 2003).
11. Tans, S. J., Verschuere, R. M. & Dekker, C. Room temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
12. Javel, A. *et al.* High-k dielectrics for advanced carbon-nanotube transistors and logic gates. *Nature Mater.* **1**, 241–246 (2002).
13. Rosenblatt, S. *et al.* High-performance electrolyte gated carbon nanotube transistors. *Nano Lett.* **2**, 869–872 (2002).
14. Cui, Y., Lauhon, L. J., Gudiksen, M. S., Wang, J. & Lieber, C. M. Diameter-controlled synthesis of single crystal silicon nanowires. *Appl. Phys. Lett.* **78**, 2214–2216 (2001).
15. Huang, Y., Duan, X., Wei, Q. & Lieber, C. M. Directed assembly of one dimensional nanostructures into functional networks. *Science* **291**, 630–633 (2001).
16. Tao, A. *et al.* Langmuir-Blodgett silver nanowire monolayers for molecular sensing using surface-enhanced Raman spectroscopy. *Nano Lett.* DOI:10.1021/nl344209 (published online 02 August 2003).
17. Whang, D., Jin, S., Wu, Y. & Lieber, C. M. Large-scale hierarchical organization of nanowire arrays for integrated nanosystems. *Nano Lett.* DOI:10.1021/nl0345062 (published online 05 August 2003).
18. Sze, S. M. *Physics of Semiconductor Devices* (Wiley, New York, 1981).
19. Hastas, N. A., Dimitriadis, C. A., Brini, J. & Kamarinos, G. Hot-carrier-induced degradation in short p-channel nonhydrogenated polysilicon thin-film transistors. *IEEE Trans. Electron Devices* **49**, 1552–1557 (2002).
20. Mizuno, T., Sugiyama, N., Kurobe, A. & Takagi, S. Advanced SOI p-MOSFET with strained-Si channel on SiGe-on-insulator substrate fabricated by SIMOX technology. *IEEE Trans. Electron Devices* **48**, 1612–1618 (2001).
21. Yeh, C., Chen, T., Gudmundsson, J. T. & Lieberman, M. A. Hydrogenation of polysilicon thin-film transistors in a planar inductive H_2/Ar discharge. *RON Device Lett.* **20**, 223–225 (1999).
22. Lauhon, L. J., Gudiksen, M. S., Wang, D. & Lieber, C. M. Epitaxial core-shell and core-multi-shell nanowire heterostructures. *Nature* **420**, 57–61 (2002).
23. Duan, X. & Lieber, C. M. General synthesis of compound semiconductor nanowires. *Adv. Mater.* **12**, 298–302 (2000).
24. Duan, X., Huang, Y., Agharwal, R. & Lieber, C. M. Single nanowire injection laser. *Nature* **421**, 241–245 (2003).
25. Weimer, P. K. The TFT-a new thin-film transistors. *Proc. IEEE* **56**, 1462–1465 (1962).
26. Madelung, O. (ed.) *Landolt-Bornstein New Series Vol. III/22a, Semiconductors: Intrinsic properties of Group IV Elements and III–V and II–VI and I–VII Compounds* 407 (Springer, Berlin, 1987).
27. Horowitz, P. & Hill, W. *The Art of Electronics* (Cambridge Univ. Press, Cambridge, 1989).
28. Duan, X., Huang, Y., Wang, J., Cui, Y. & Lieber, C. M. Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices. *Nature* **409**, 66–69 (2001).
29. Xia, Y., Qin, D. & Whitesides, G. M. Microcontact printing with a cylindrical rolling stamp: A practical step toward automatic manufacturing of patterns with submicrometer-sized features. *Adv. Mater.* **8**, 1015–1017 (1996).
30. Sirringhaus, H. *et al.* High-resolution inkjet printing of all-polymer transistor circuits. *Science* **290**, 2123–2126 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. Bock and C. Chow for insights and support; D. Stumbo for discussions and opinions on the manuscript; Y. Pan and H. Liu for supplying Si/SiN_x wafers; and C. M. Lieber, H. Park and P. McEuen for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.D. (xduan@nanosysinc.com).

High CO₂ levels in the Proterozoic atmosphere estimated from analyses of individual microfossils

Alan J. Kaufman¹ & Shuhai Xiao^{2*}

¹Geology Department, University of Maryland, College Park, Maryland 20742-4211, USA

²Department of Earth and Environmental Sciences, Tulane University, New Orleans, Louisiana 70118, USA

* Present address: Department of Geological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA

Solar luminosity on the early Earth was significantly lower than today. Therefore, solar luminosity models suggest that, in the atmosphere of the early Earth, the concentration of greenhouse gases such as carbon dioxide and methane must have been much higher^{1,2}. However, empirical estimates of Proterozoic levels of atmospheric carbon dioxide concentrations have not hitherto been available. Here we present ion microprobe analyses of the carbon isotopes in individual organic-walled microfossils extracted from a Proterozoic (~1.4-gigayear-old) shale in North China. Calculated magnitudes of the carbon isotope fractionation in these large, morphologically complex microfossils suggest elevated levels of carbon dioxide in the ancient atmosphere—between 10 and 200 times the present atmospheric level. Our results indicate that carbon dioxide was an important greenhouse gas during periods of lower solar luminosity, probably dominating over methane after the atmosphere and hydrosphere became pervasively oxygenated between 2 and 2.2 gigayears ago.

The organic remains used in this study consisted of a single acritarch species (*Dictyosphaera delicata*, Fig. 1a, b) from the Ruyang Group in Shanxi Province³. An intrusive granite radiometrically constrains the succession to be >999 Myr old. Although stratigraphic variations in ¹³C abundance throughout the Ruyang are consistent with marine sediments deposited between about 1,250 and 1,900 Myr ago⁴, the equivalence of microfossil assemblages in the Ruyang and the radiometrically constrained Roper Group of Australia suggests an age closer to 1,400 Myr (ref. 5). Acritarchs of this antiquity are interpreted as resting cysts of photosynthetic eukaryotes, although their organic encasements were synthesized during active growth and should thus reflect primary photosynthetic processes. *D. delicata* is characterized by a vesicle (150 ± 40 μm in diameter) consisting of interlocked polygonal plates (1–3 μm). Its large size and complex morphology suggest that it was probably a photosynthetic eukaryote and, by inference, used the Calvin cycle for carbon fixation⁶. Although similarly sized sulphur-oxidizing bacteria have been found in restricted modern environments⁷, these are morphologically simple prokaryotes lacking a robust encystment stage and are unlikely to be preserved in Proterozoic-aged shales.

To evaluate potential surface contamination of the carbonaceous acritarchs by bacteria or acid-resistant minerals, the fossils were observed at nanometre resolution using a field-emission scanning electron microscope (FE-SEM). A handful of 1–4-μm-long rod-like structures were observed on the surface of the microfossil (Fig. 1c), but they may have resulted from surface folding of the acritarch walls; whatever their origin, their low abundance makes it unlikely that they represented a significant carbon contamination for the ion microprobe analyses. Close examinations of the 1–3-μm polygonal plates that characterize the microfossil revealed a patterned surface of high and low electron density (Fig. 1d); the areas of low density appear to be pores of 50–100 nm diameter. Interior surfaces were

imaged through a pustular opening on the surface of one of the specimens (Fig. 1e). The network of surface pores was also imaged in the interior, but in this case the pores appeared filled with a material of higher electron density, in strong contrast to the carbonaceous walls that were uniform throughout (Fig. 1f). The preservation of the mineral cement on the interior suggests that it was isolated from the acid environment during sample preparation with hydrofluoric acid; on and near the surface, the pore-filling cement was probably removed during acidification. To determine the mineralogical composition of this cement, ultra-thin sections of the specimens were analysed using a transmission electron microscope (TEM). The cement is largely composed of an unusual aluminium phosphate mineral (florencite group) containing 1–2% Sr and up to 5% rare-earth elements. Elemental maps confirm that the exterior surface is composed of more than 90% carbon; isolated 'hot spots' with elevated Ca, O and P occur at tears in the cell wall surface, suggesting the presence of phosphate minerals in the interior.

Carbon-isotopic compositions of the microfossils were measured using a Cameca 6f ion microprobe at high mass resolution. The magnitude of instrumental mass fractionation (IMF) in secondary ion mass spectrometry is probably a function of the mean atomic weight of the analysed surface, as well as topographic irregularities⁸. In this regard, the absence of the phosphate cement on the surface and in the ion microprobe pits was fortuitous. Six individual *D. delicata* specimens were each analysed multiple times ($n = 27$ analyses reported in the standard δ notation; Table 1). Given the micrometre-scale relief of the microfossils relative to the highly polished Mao diamond (used as a standard; see Methods), the 1.5‰ precision of fossil carbon isotope analyses suggests that the topographic effect on IMF was small. The averaged carbon-isotopic compositions of individuals ranged narrowly between –33.4‰ and

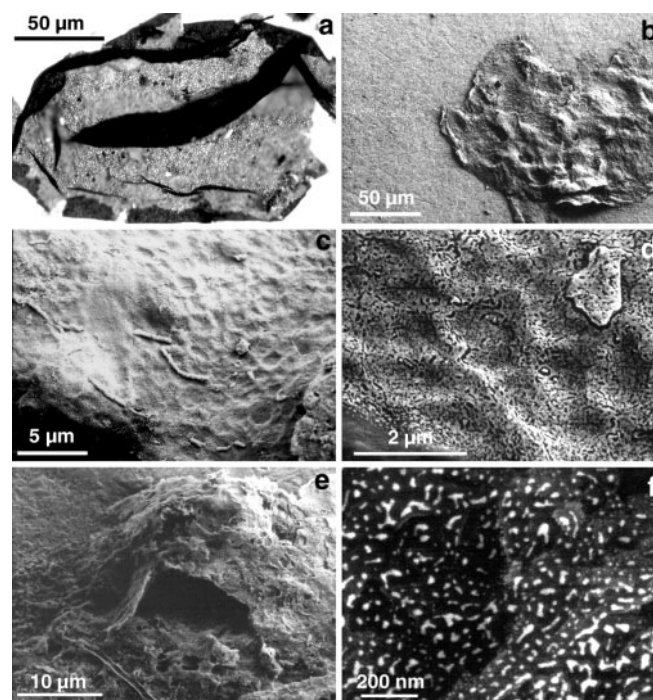


Figure 1 Photographs of *D. delicata* specimens used in this study. **a**, Polygonal pattern on *D. delicata* vesicle wall seen with transmitted light. **b–f**, FE-SEM images of compressed *D. delicata* (**b**, right) mounted on highly polished aluminium plate (**b**, left), vesicle wall showing polygonal pattern and rod-like structures (**c**), nanoscale pores (lower electron density) on outer vesicle surface (**d**), a pustular opening through which inner vesicle surface was imaged (**e**), and nanoscale pores on interior vesicle surface filled with a material of higher electron density (**f**).

Table 1 IMF corrected carbon isotope values

Specimen*	Number	Individual $\delta^{13}\text{C}$ values (‰, V-PDB)	Mean $\delta^{13}\text{C}$ value (‰, V-PDB)	Standard deviation (1 σ , ‰)
A	5	−31.9; −34.6; −35.2; −32.5; −32.7	−33.4	1.6
B	5	−36.7; −34.0; −33.2; −35.3; −33.7	−34.6	1.4
F	2	−36.8; −35.9	−36.4	0.6
I	4	−31.5; −35.1; −33.5; −34.0	−33.5	1.5
J	5	−35.8; −33.7; −35.1; −36.2; −37.8	−35.7	1.5
K	6	−33.9; −36.7; −38.4; −37.9; −33.0; −34.6	−35.7	2.2
Kerogen†	5	−28.9; −29.5; −33.4; −36.4; −31.5	−31.9	3.1

**D. delicata* specimens.

†Kerogen isolate (see text).

$\delta^{13}\text{C} = \left[\left(\frac{^{13}\text{C}/^{12}\text{C}_{\text{sample}}}{^{13}\text{C}/^{12}\text{C}_{\text{V-PDB}}} - 1 \right) \times 1,000 \right]$

−36.4‰, with an overall species average of −34.9‰. The pooled data have a similar average (−34.8‰) but larger range (−31.5‰ to −38.4‰), probably reflecting both topographic effects and isotopic differences between specimens. These microfossils are 3–6‰ more depleted in ^{13}C than kerogen isolated from the fossil-bearing shale, which has been measured by ion microprobe at -31.9 ± 3.1 ‰ ($n = 5$; the larger variance is probably related to kerogen heterogeneity and topographic irregularities on the pressed pellet) and by conventional techniques at -30.1 ± 0.4 ‰.

These analyses represent an improvement on previous ion microprobe analyses of individual Precambrian microfossils^{9,10}. The choice of large species is more appropriate to the size of the ion microprobe beam, allowing for multiple measurements of individuals. In addition, acid extraction isolated and cleaned microfossil surfaces of the phosphate cement, which could have caused IMF variations. Insofar as kerogen extracted from ancient

sediments typically contains both primary and secondary inputs of organic matter, these microfossil analyses are interpreted to more closely reflect the isotopic composition of primary photosynthate.

Thus, coupled with measurement of well-preserved marine carbonate, these ion microprobe $\delta^{13}\text{C}$ analyses allow magnitudes of carbon isotope fractionation (ϵ_p)—related to specific carbon fixation pathways—to be determined. Given that bedded carbonates closely associated with the fossiliferous Ruyang shales have $\delta^{13}\text{C}$ values of near −1‰ (ref. 3), the dissolved carbon dioxide that served as a food source to these photoautotrophs would have values ranging between −8‰ and −10‰, assuming surface temperatures between 15 and 30°C. Therefore, calculated ϵ_p values for our specimens of *D. delicata* range between 23.4‰ and 28.4‰ (see Methods). The average value is ~25‰, which is the maximum known fractionation recorded by organisms that use Calvin cycle metabolism (Fig. 2). This biochemical assignment is consistent with the morphological complexity of *D. delicata*, suggesting that the Calvin cycle was in place at least 1,400 Myr ago¹¹.

By quantifying ϵ_p the ion microprobe technique provides an empirical constraint of palaeo- p_{CO_2} levels. Chemostat studies of modern eukaryotic algae show that ϵ_p in the Calvin cycle is controlled by dissolved CO_2 concentrations as well as the availability of micronutrients, growth rate, and the volume-to-surface (V/S) ratio^{12–15}. Modern phytoplankton have a V/S near 1 μm , considering these organisms as simple geometric shapes lacking external processes. Maximal ϵ_p of ~25‰ for modern protists is attained at CO_2 concentrations of 8–10 times the present atmospheric level (PAL; Fig. 2). If the same CO_2 palaeobarometer is applied to the Ruyang sediments, consideration must be given to the enormous size of the microfossils. Based on a simple 150- μm -diameter sphere, *D. delicata* would have a V/S of 25 μm , which results in calculated concentrations of dissolved CO_2 in sea water (C_e) two orders of magnitude greater than present-day values (~10 $\mu\text{mol kg}^{-1}$) and a atmospheric p_{CO_2} of >200 times PAL, depending on the growth rate and temperature, which remains uncertain. However, the effective V/S of *D. delicata* might have been less if it had been ornamented with processes similar to *Shuiyousphaeridium macroreticulatum*, a similarly sized acritarch in the same assemblage³. It is possible that our specimens of *D. delicata* have lost their processes during diagenesis¹⁶ or sample processing. If these spines were part of the photosynthetic apparatus they would reduce the effective V/S ; modelled with 1,000 of these spines, V/S drops to near that of modern phytoplankton, and hence an estimated p_{CO_2} of >10 times PAL.

The empirically estimated p_{CO_2} levels can be compared with model results. Calculations of atmospheric p_{CO_2} levels required to compensate for reduced solar luminosity in the Proterozoic (2,500–543 Myr ago) range widely¹, but may have been as high as 1,000 times PAL (Fig. 3). This is consistent with predictions of higher seawater alkalinity (buffering capacity for the excess dissolved CO_2) inferred from the predominance of carbonate cements in the ancient record¹⁷. If the ocean's buffering capacity were exceeded the carbonate compensation depth would rise, and eventually no carbonates would form. The limit of CO_2 that could be absorbed

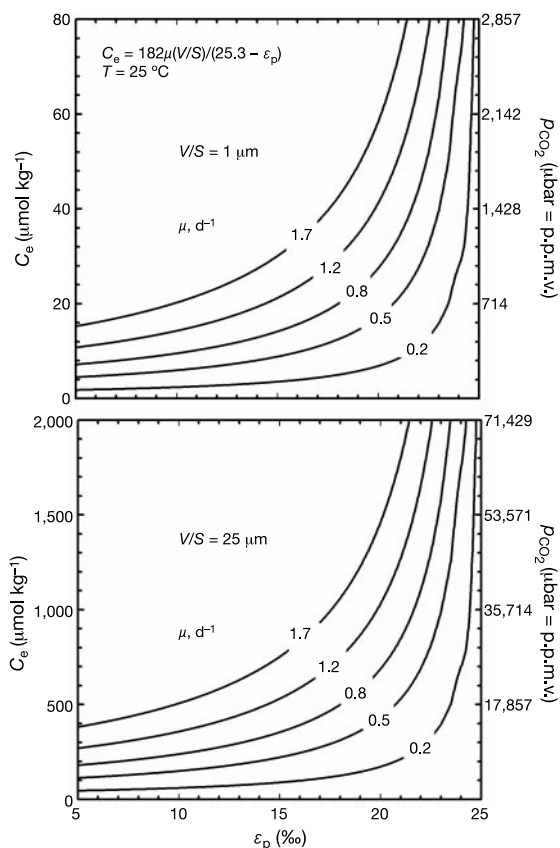


Figure 2 Relationship between carbon isotopic fractionation of the Calvin cycle (ϵ_p) and CO_2 concentration in sea water (C_e) at different growth rates (μ , in units of d^{-1}). Top, for modern phytoplankton with volume-to-surface ratio (V/S) of 1 μm ; bottom, for *D. delicata* with V/S of 25 μm . Based on experimental data^{12–15}.

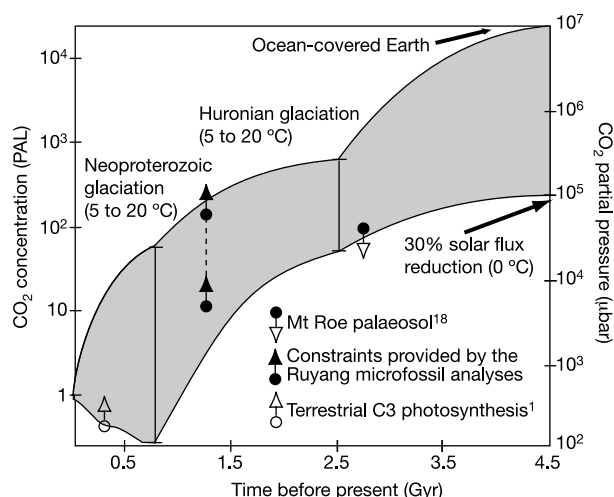


Figure 3 Atmospheric p_{CO_2} levels in the Archaean and Proterozoic. Shaded area, modelled in ref. 1. Empirical constraints for Archaean (from Mt Roe palaeosol)¹⁸ and middle Proterozoic (this study) atmospheric p_{CO_2} levels are indicated. The two connected upward-pointing arrows at 1.4 Gyr ago relate to p_{CO_2} estimates based on the possible V/S of *D. delicata* (see text). After ref. 1.

and still allow shallow water carbonate deposition (required by the geological record) is likely to be less than 100 times PAL. However, an absolute estimate of this partial pressure cannot be made without independent assumptions of the size and alkalinity of the ancient oceans.

The atmospheric p_{CO_2} range (>10–200 times PAL) determined by this ion microprobe study is broadly consistent with concentrations required to compensate for lower Proterozoic solar luminosity (Fig. 3). It is, however, generally higher than the constraint provided by the stability of siderite in a 2,700-million-year-old palaeosol from Western Australia¹⁸, which requires CH_4 —a compensatory greenhouse gas—to have been relatively more abundant in the Archaean atmosphere. After the atmosphere became pervasively oxygenated and CH_4 concentrations decreased some 2.0–2.2 billion years ago^{1,19–21}, elevated CO_2 concentrations alone could have compensated for the lower heat flux from the Sun for the remainder of the Proterozoic eon and beyond. This empirical estimate, if confirmed by ion microprobe analysis of similarly aged organic-walled microfossils, places an important constraint on geochemical models of CH_4 concentrations in the Proterozoic atmosphere (compare ref. 2).

Methods

Acritarch isolation

Shale samples were collected from the Beidajian Formation of the Ruyang Group at Shuiyongou, Shanxi Province, North China. About 100 g of fresh sample was reacted with concentrated HF for three days and then rinsed repeatedly with distilled water. Taxonomically identified acritarchs and bulk kerogen were physically separated under a binocular microscope.

Ion microprobe

Specimens of *D. delicata* were mounted on a highly polished aluminium plate in a holder containing the standard Mao diamond. Carbon-isotopic compositions of the microfossils were measured with the Carnegie Institution Department of Terrestrial Magnetism Cameca 6f ion microprobe. The impact energy of the primary Cs^+ beam includes the acceleration energy at the source (10 keV) plus a further acceleration as the beam approaches the sample at -5 keV electric potential. Beam intensity was 0.1–0.3 nA and was focused down to a 20–25- μm spot, which allowed for multiple analyses of the same individual. Crater depths are estimated to be ~ 2 – 5 μm . Negatively charged carbon ions were admitted into the mass spectrometer, which counted 100 cycles after a five-minute pre-sputter treatment. Secondary ions were counted with an ETP electron multiplier. With a thin gold coat on the fossils, over 10^6 counts per second were recorded on the major ^{12}C

secondary beam during analyses of both the standard Mao diamond and microfossils. The ^{12}C hydride peak was clearly resolved from those of ^{13}C ions and did not interfere in the isotopic measurement²².

An IMF correction factor was determined by repeated measurements of the standard Mao diamond ($\delta^{13}\text{C} \approx -6.5\text{‰}$ versus V-PDB determined by conventional techniques)²². Acritarchs were measured intermittently over three consecutive days, during which the standard showed a high degree of external reproducibility ($\delta^{13}\text{C}_{\text{IMF uncorrected}} = -47.4 \pm 0.3\text{‰}$; $n = 8$; IMF $\approx 40.9\text{‰}$). A pressed pellet of organic matter (kerogen) extracted from the fossiliferous shale was measured in a separate single-day session, during which the standard was measured at $\delta^{13}\text{C}_{\text{IMF uncorrected}} = -45.1 \pm 0.1\text{‰}$ ($n = 2$; IMF $\approx 38.6\text{‰}$). Earlier measurements made with this Cameca 6f of both crystalline graphite and diamond show that they exhibit very similar IMF²², although the IMF of finely ground graphite was significantly different, probably due to a scattering effect associated with the size of individual grains relative to the Cs^+ ion beam.

Calculations of ϵ_p and p_{CO_2}

Dissolved CO_2 concentrations were calculated using a model presented in Fig. 2 (refs 12–15). To avoid intra-specimen topographic effects, we used the averaged $\delta^{13}\text{C}$ values for each specimen in the calculation of ϵ_p values. The range of ϵ_p values for different specimens (between 23.4‰ and 28.4‰) may relate to temperature effects on the $\delta^{13}\text{C}$ of dissolved inorganic carbon and carbonate, used to determine ϵ_p from ancient sediments¹¹. Those fossils with ϵ_p values of $>25\text{‰}$ (the experimentally determined maximum value) may have grown during seasons with cooler averaged surface temperature, or these may represent specimens that reveal some degree of taphonomic bias. Atmospheric p_{CO_2} levels were calculated assuming ocean-atmosphere equilibrium; this is a reasonable assumption given that *D. delicata* lived in the photic zone. The sensitivity of the estimated p_{CO_2} levels depends critically on uncertainties of surface temperature and V/S .

FE-SEM and elemental mapping

Specimens of *D. delicata* on the highly polished aluminium plate were imaged using the FE-SEM at the University of Maryland Laboratory for Ion Beam Research and Applications. Magnifications up to 200,000 $\times$ were achieved during the investigation and structures as small as 5 nm were resolvable. The specimen was then observed using the University of Maryland JXA-8900 Superprobe, a high resolution SEM and WDS/EDS combined electron probe microanalyser. Elemental (C, O, Ca, P) maps were created for one of the specimens.

TEM

Individual acritarchs were embedded in epoxy and 70-nm sections were prepared at the Tulane University Coordinated Instrumentation Facility. The sections were observed with the Johns Hopkins University Philips 420 TEM operating at 120 kV. X-ray spectra were collected using an Oxford Instruments light-element energy dispersive X-ray analyser. Semi-quantitative analyses were obtained using the DTSA program interfacing with a 4Pi multichannel analyser. Data were fitted by filtered reference spectra and reduced following the assumptions of the Cliff-Lorimer thin-film analysis method and empirical k -factors.

Received 22 January; accepted 10 August 2003; doi:10.1038/nature01902.

- Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
- Pavlov, A. A., Hurtgen, M. T., Kasting, J. F. & Arthur, M. A. Methane-rich Proterozoic atmosphere? *Geology* **31**, 87–90 (2003).
- Xiao, S., Knoll, A. H., Kaufman, A. J., Yin, L. & Zhang, Y. Neoproterozoic fossils in Mesoproterozoic rocks? Chemostratigraphic resolution of a biostratigraphic conundrum from the North China Platform. *Precamb. Res.* **84**, 197–220 (1997).
- Brasier, M. D. & Lindsay, J. F. A billion years of environmental stability and the emergence of eukaryotes: New data from northern Australia. *Geology* **26**, 555–558 (1998).
- Javaux, E. J., Knoll, A. H. & Walter, M. R. Morphological and ecological complexity in early eukaryotic ecosystems. *Nature* **412**, 66–69 (2001).
- Shively, J. M. & Barton, L. L. *Variations in Autotrophic Life* 1–346 (Academic, New York, 1991).
- Schulz, H. et al. Dense populations of a giant sulfur bacterium in Namibian shelf sediments. *Science* **284**, 493–495 (1999).
- Cristy, S. S. in *Inorganic Mass Spectrometry* (eds Barshick, C.M., Duckworth, D.C. & Smith, D.H.) 159–221 (Marcel Dekker, New York, 2000).
- House, C. H. et al. Carbon isotope composition of individual Precambrian microfossils. *Geology* **28**, 707–710 (2000).
- Ueno, Y., Isozaki, Y., Yurimoto, H. & Maruyama, S. Carbon isotopic signatures of individual Archean microfossils(?) from Western Australia. *Int. Geol. Rev.* **43**, 196–212 (2001).
- Hayes, J. M., Strauss, H. & Kaufman, A. J. The abundance of C-13 in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **16**, 103–125 (1999).
- Laws, E. A., Popp, B. N., Bidigare, R. R., Kennicutt, M. C. & Macko, S. A. Dependence of phytoplankton carbon isotopic composition on growth rate and $[\text{CO}_2(\text{aq})]$: Theoretical considerations and experimental results. *Geochim. Cosmochim. Acta* **59**, 1131–1138 (1995).
- Laws, E. A., Bidigare, R. R. & Popp, B. N. Effect of growth rate and CO_2 concentration on carbon isotopic fractionation by the marine diatom *Phaeodactylum tricornutum*. *Limnol. Oceanogr.* **42**, 1552–1560 (1997).
- Popp, B. N. et al. Effect of phytoplankton geometry on carbon isotopic fractionation. *Geochim. Cosmochim. Acta* **62**, 69–77 (1998).
- Bidigare, R. R. et al. Consistent fractionation of C-13 in nature and in the laboratory: Growth-rate effects in some haptophyte algae. *Glob. Geochem. Cycles* **13**, 251–252 (1999).
- Knoll, A. H. Microbiotas of the late Precambrian Hunsberg Formation, Nordaustlandet, Svalbard. *J. Paleontol.* **58**, 131–162 (1984).

17. Grotzinger, J. P. & Kasting, J. F. New constraints on Precambrian ocean composition. *J. Geol.* **101**, 235–243 (1993).
18. Rye, K., Kuo, P. H. & Holland, H. D. Atmospheric carbon-dioxide concentrations before 2.2-billion years ago. *Nature* **378**, 603–605 (1995).
19. Holland, H. D. in *Early Life on Earth* (ed. Bengtson, S.) 237–244 (Columbia Univ. Press, New York, 1994).
20. Karhu, J. & Holland, H. Carbon isotopes and the rise of atmospheric oxygen. *Geology* **24**, 867–870 (1996).
21. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the Archean oceans. *Science* **298**, 2372–2374 (2002).
22. Farquhar, J., Hauri, E. & Wang, J. New insights into carbon fluid chemistry and graphite precipitation: SIMS analysis of granulite facies graphite from Ponnudi, South India. *Earth Planet. Sci. Lett.* **171**, 607–621 (1999).

Acknowledgements We thank the following for technical assistance: E. Hauri, J. Wang, J. Orloff, R. Dotson, K. Livi, D. Veblen and P. Piccoli. We also thank A. Knoll and L. Yin for providing samples for bulk rock analyses, and J. Hayes for comments on an earlier version of this manuscript. This research was supported by NASA Exobiology, NSF Geology and Paleontology, and China MOST 973 programs.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.K. (kaufman@geol.umd.edu) or S.X. (xiaoxiao@vt.edu).

Fragments of the earliest land plants

Charles H. Wellman¹, Peter L. Osterloff^{2*} & Uzma Mohiuddin²

¹Department of Animal and Plant Sciences, University of Sheffield, Alfred Denny Building, Western Bank, Sheffield S10 2TN, UK

²Petroleum Development Oman, PO Box 81, Muscat 113, Sultanate of Oman

* Present address: Sarawak Shell Berhad, Locked Bag No.1, 98009 Miri, Sarawak, Malaysia

The earliest fossil evidence for land plants comes from microscopic dispersed spores^{1–3}. These microfossils are abundant and widely distributed in sediments, and the earliest generally accepted reports are from rocks of mid-Ordovician age (Llanvirn, 475 million years ago)⁴. Although distribution, morphology and ultrastructure of the spores indicate that they are derived from terrestrial plants, possibly early relatives of the bryophytes, this interpretation remains controversial⁵ as there is little in the way of direct evidence for the parent plants. An additional complicating factor is that there is a significant hiatus between the appearance of the first dispersed spores and fossils of relatively complete land plants (megafossils)⁶: spores predate the earliest megafossils (Late Silurian, 425 million year ago) by some 50 million years⁷. Here we report the description of spore-containing plant fragments from Ordovician rocks of Oman. These fossils provide direct evidence for the nature of the spore-producing plants. They confirm that the earliest spores developed in large numbers within sporangia, providing strong evidence that they are the fossilized remains of bona fide land plants. Furthermore, analysis of spore wall ultrastructure supports liverwort affinities.

During the Ordovician period Oman lay on the northwest margin of the continent Gondwana, at mid-latitudes to the south of the palaeoequator. Borehole Ghaba-1 (northern Oman) penetrates a thick Ordovician sequence, consisting primarily of terrestrial and marginal marine deposits, but punctuated by conspicuous shale units representing rare but significant marine incursions⁸. The fossils described herein were recovered from a monotonous sequence of grey-green sandy siltstones and fine sandstones (Ghaba-1: core 21, 1542.0–1545.3 m) belonging to the Hasirah Member (Safiq Formation: Haima Supergroup). This member comprises a sequence of shales deposited during a major marine transgression, overlain by coarser sediments interpreted as marginal

marine and terrestrial deposits that accumulated during a continuously regressive phase that followed^{8–9}. The fossils derive from the latter and are considered to be Late Ordovician (Caradoc) in age based on palynology⁹ and sequence stratigraphy⁸. The fossil-bearing deposits yield non-marine palynomorph assemblages indicative of a Caradoc age. They are further constrained because they are sandwiched between distinct and regionally continuous shale units, representing the discrete marine incursions, which yield rich *in situ* marine palynomorph assemblages that are reliably dated⁹, and are correlated with surface exposures dated using marine invertebrate megafossils⁸.

Thirteen samples from core 21 were prepared by standard palynological acid maceration techniques, and sieved using a 10- μ m mesh. The palynological residues are dominated by well preserved and thermally immature spores, although some samples contained rare marine elements (acritarchs, chitinozoans, scolecodonts). These data suggest a marginal marine setting, a predominantly non-marine environment with minor marine incursions, or an entirely non-marine sequence into which the rare marine elements are reworked.

Large quantities of three of the samples were top-sieved using 100- μ m mesh in the hope of retrieving larger fragments of the spore producers. This yielded a total of 12 large (0.24–0.49 mm) plant fragments. These were examined using scanning electron microscopy (SEM) and then sectioned for transmission electron microscopy (TEM) analysis. The fragments consist of spore masses, some of which preserve part of the presumed sporangium covering. Some are disc-shaped with a relatively large area of covering preserved, and probably represent relatively complete contents of originally spherical sporangia. Individual spore masses all contain identical spores comprising naked tetrads, envelope-enclosed tetrads or naked dyads. Presumably these spore units were ‘permanent’ (that is, did not separate before dispersal) because separating tetrads and dyads are unknown before the latest Ordovician³. Late Silurian coprolites, consisting predominantly of spores, are similar in size and appearance to the described fossils¹⁰; however, they are easily distinguished because the coprolites are elongated, do not possess any form of covering, and usually contain a number of different spore types in addition to fragments of plant debris, such as cuticle.

Figure 1a–e illustrates a specimen containing naked permanent tetrads. The fragment measures 0.30 by 0.23 mm and it is estimated that it contains approximately 2,620 tetrads (that is, 10,480 single spores). The spore mass overlies part of the presumed sporangium covering (Fig. 1a arrow, c), which is 3.6 μ m in thickness and appears entirely featureless on the surface and in cross-section. The specimen illustrated in Fig. 1f, g is 0.30 mm in diameter and contains an estimated 2,670 naked permanent tetrads. Here, however, a much greater area of covering (about 5,680 μ m²) is preserved. Again it is thin (about 2 μ m) and entirely featureless. Figure 1h illustrates an envelope-enclosed permanent tetrad preserved in another sporangium. The envelope is clearly discernible and has a sculpture of fine muri.

TEM analysis of sections of the fragments revealed details of spore wall ultrastructure, in addition to evidence pertaining to the nature of the covering/spore mass contact. In two specimens, comprising naked permanent tetrads, the spore walls are entirely homogeneous. In one specimen (Fig. 2), comprising naked permanent dyads, exquisite spore wall ultrastructure is preserved. The walls are 0.72–1.01 μ m in thickness. The bulk of the wall consists of a stacked series of continuous, parallel lamellae. The lamellae are a consistent 41–58 nm in thickness, and up to 11 lamellae are discernible in each wall. However, the lamellae become less clearly defined towards the inside and outside of the wall, and in some spores these parts appear homogeneous. It is likely that the entire wall was originally lamellated, but the lamellae have been obscured by sporopollenin accretion. Preservation of wall ultrastructure is fickle, and is dependent on both ontogenetic stage and subtle diagenetic effects. Although it

is likely that there is a genuine difference between homogeneous walls of naked permanent tetrads and lamellate wall of naked permanent dyads¹¹, these differences may relate to ontogeny and/or diagenesis.

Sections through the presumed sporangial covering were observed only in the specimen illustrated in Fig. 1a–e. The covering clearly overlies the outermost spores, and in fact penetrates into the gaps between them. The section of preserved covering is 1.6 μm in thickness. Unlike the spore walls that are homogeneous, it is extremely variable in nature. It is made up of convoluted ‘swirls’ of material, distinguishable because of their differing electron density. Also there are some small, irregularly shaped voids. The covering becomes less ‘swirly’ and more regular towards the outside. The swirly material possibly represents sporopollenin released as the tapetum lining the sporangial wall degenerated. Similar sporangium wall linings, interpreted as a residue of the tapetum, are reported in a number of extant plant groups¹², in addition to fossils of early land plants¹³.

Ordovician dispersed spore assemblages consist of spores of unusual morphology that are termed cryptospores¹⁴ (permanent tetrads, permanent dyads and monads that are either naked or enclosed within an envelope). They have been reported from throughout the globe, exhibiting surprisingly little temporal and spatial variation^{3,15}, and this uniformity suggests a worldwide cosmopolitan flora. The Oman dispersed spore assemblages provide further evidence for this hypothesis, as they are identical to coeval assemblages from elsewhere.

There are several lines of evidence indicating that Ordovician cryptospores derive from land plants^{2–3}. First, the distribution of the spores is similar to that of extant spores/pollen (that is, they occur in non-marine sediments, and when found in marine sediments their abundance declines offshore). Second, they are similar in size and

morphology to the spores of land plants. In fact, some of the permanent tetrads, including envelope-enclosed forms, closely resemble the spores of extant liverworts². Third, wall ultrastructure in Ordovician cryptospores is varied, but in some morphotypes suggests affinities with extant liverworts¹⁶. Fourth, and rather intriguingly, some of the earliest known land plant mesofossils/megafossils from the latest Silurian to earliest Devonian contain spores similar in morphology to the Ordovician forms (for example, permanent tetrads and dyads)^{17–20}. These cryptospore producers are of uncertain affinity, although some possess tantalizing bryophyte-like characteristics¹⁷, and it is possible that they represent relic elements of the flora.

The *in situ* spores described herein are clearly identical to coeval dispersed forms proposed as the earliest evidence for land plants. The new fossils confirm that the spores were produced in large numbers in sporangia, providing further, convincing evidence that they derive from land plants. The ability to produce abundant sporopollenin-coated spores is considered to be an adaptation vital for colonization of the terrestrial environment²¹. It is clear that the sporangia were extremely small, but this is not unexpected. Sporangia of extant bryophytes are rarely more than 1 mm in diameter. It is calculated that a spherical sporangium of this size could contain 95,000 spore tetrads of similar size to the Ordovician forms. Indeed some of the Oman fossils contain a minimum of 7,450 tetrads. In recent years our search image for the earliest plants has begun to focus on very small plants, based on the small size of extant bryophytes and diminutive cryptospore producers recovered from the Late Silurian to Early Devonian (the so-called Lilliputian plants¹³). Clearly our discovery of fragments of very early plants, which suggest that the plants were miniscule, supports such an hypothesis.

Spore wall development is conserved and variable between

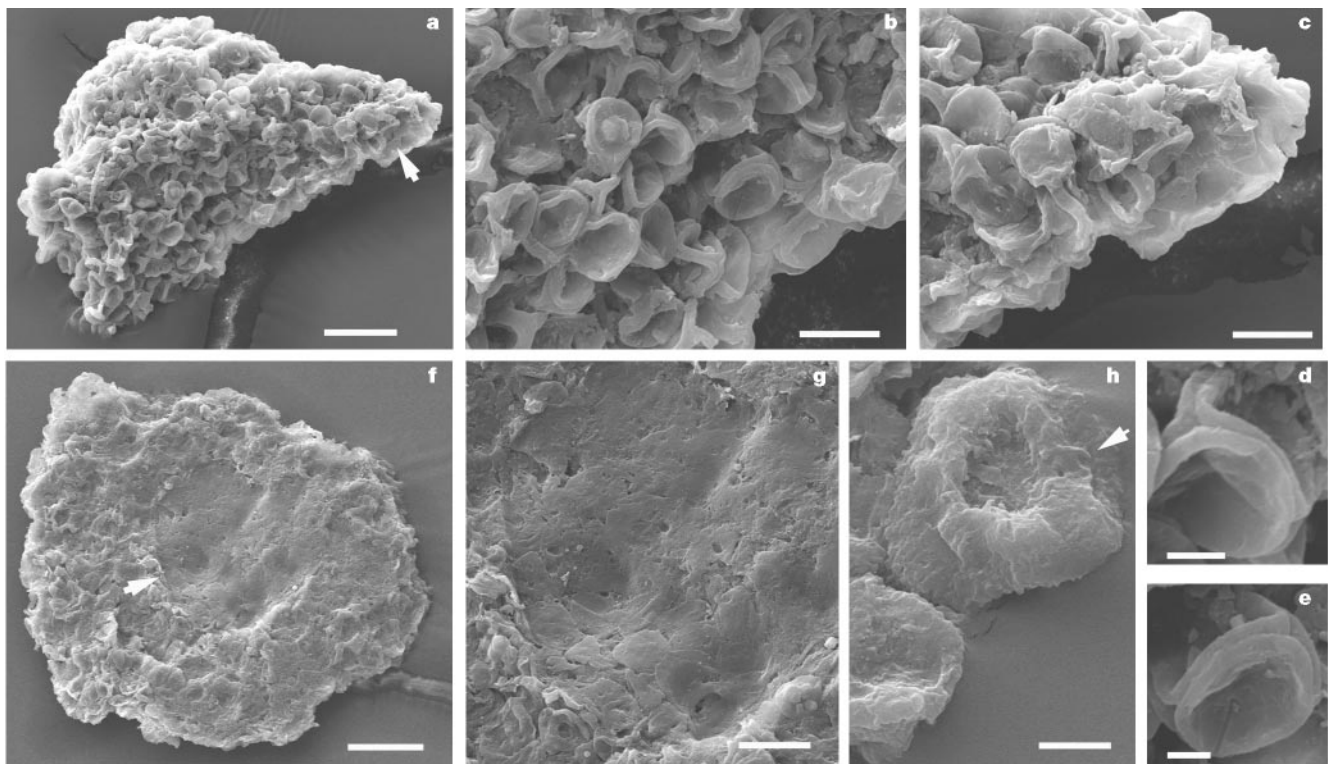


Figure 1 Fossil plant fragments from borehole Ghaba-1, core 21. **a**, Specimen CW47a. SEM of fragment of sporangium containing naked permanent tetrads. Note the presence of sporangium covering in the bottom right-hand corner (arrow). **b**, Close-up of **a** illustrating the spore contents. **c**, Close-up of **a** illustrating spores overlying the sporangium covering. **d, e**, Close-up of **a** illustrating individual spore tetrads. **f**, Specimen

CW47f. SEM of relatively complete sporangium, with a large patch of sporangium covering preserved (arrow). **g**, Close-up of **f** illustrating the nature of the sporangium covering. **h**, Specimen CW47i. SEM of an envelope-enclosed permanent tetrad that is preserved in a fragmentary sporangium. Note the muri ornamenting the envelope (arrow). Scale bars: **a**, 50 μm ; **b, c**, 20 μm ; **d, e**, 5 μm ; **f**, 75 μm ; **g**, 30 μm ; **h**, 10 μm .

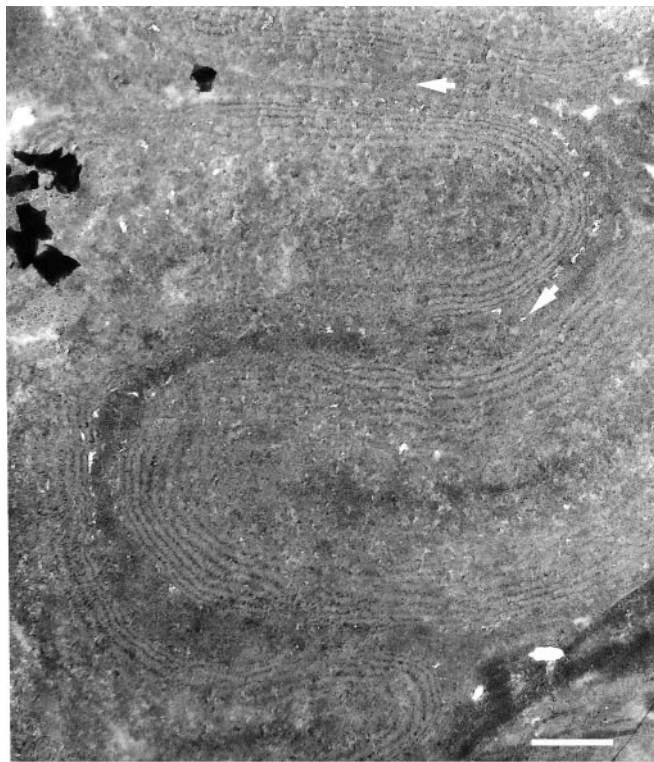


Figure 2 Specimen CW47c. TEM of part of a sporangium illustrating partial sections through three spores. The top arrow marks contact between the upper spore and the sporangium wall. The bottom arrow marks contact between the spores to the lower left and lower right. Note the lamellate ultrastructure. Scale bar, 0.5 μm .

different plant groups, and hence the presence of exquisitely preserved spore wall ultrastructure provides further strong evidence of the affinities of these Ordovician plants. Lamellae are considered to be produced at some point during the ontogeny of virtually all land plant spores/pollen, and may be present in a variety of forms and positions in many mature spore/pollen walls. However, spores/pollen in which continuous, parallel-arranged lamellae are present at maturity throughout the walls are confined to extant liverworts^{16,22}, suggesting that the Ordovician plants from Oman may have liverwort affinities. This reinforces claims for liverwort affinities based on morphology of Ordovician spores, and conforms to land plant phylogenies that place liverworts as basal.

Recent phylogenetic analyses of land plants use morphological and/or molecular evidence, and although most are based solely on extant plants, some also incorporate evidence from fossils²³. These analyses provide strong evidence that land plants are monophyletic, and that extant charophyte green algae are their sister group (that is, the earliest land plants evolved from aquatic green algal ancestors)²¹. It is generally agreed that the earliest divergent land plants were bryophytes, but there is some disagreement as to the exact relationship between the three extant bryophyte groups (liverworts, hornworts, mosses) and vascular plants. Most recent analyses indicate that the bryophytes are paraphyletic with respect to the vascular plants, with a moss/vascular plant sister group relationship, and either the liverworts or hornworts as the most primitive, early divergent, extant land plants^{23–26}.

If liverwort affinities are accepted for the Ordovician fossils, one unexpected finding is the absence of elaters. Elaters are elongate cells that occur interspersed with spores and facilitate their dispersal by hygroscopic movements. Among extant plants they occur in the liverworts and hornworts, but are absent from more derived members of the plant kingdom, possibly because the latter have evolved increasingly sophisticated sporangia with more effective

spore dispersal mechanisms²⁴. On the basis of recent phylogenies suggesting that either liverworts or hornworts are basal, and evidence that the earliest plants were liverwort-like, one might expect the earliest land plants to possess elaters. Indeed, studies of recent elaters suggest that they should survive in the fossil record²⁷. However, elaters have only rarely been reported *in situ* in the fossil record, and never as dispersed microfossils. This indicates one of the following possibilities: they were only present in certain taxa (they have been secondarily lost in many liverworts); they are a post-Caradoc adaptation; they are not readily fossilized; or they are present but not discernible (perhaps they were attached at the point where the axis meets the sporangium, as is the case in certain extant bryophytes).

The new fossils from Oman shed light on a number of controversies concerning the origin of land plants and their fossil record. They provide convincing evidence that the spores purported to provide the earliest fossil evidence for land plants do indeed derive from land plants that produced sporangia containing abundant spores. Spore wall ultrastructure also provides further evidence for liverwort affinities, although the absence of elaters is intriguing. Finally, the fossils demonstrate the diminutive size of the plants, providing a search image for further discoveries. We predict that top-sieving of palynological residues of suitable age and lithology will produce further finds of the earliest land plants. □

Received 1 April; accepted 3 July 2003; doi:10.1038/nature01884.

- Gray, J. & Boucot, A. J. Early Silurian spore tetrads from New York: earliest New World evidence for vascular plants? *Science* **173**, 918–921 (1971).
- Gray, J. The microfossil record of early land plants: advances in understanding of early terrestrialization, 1970–1984. *Phil. Trans. R. Soc. Lond. B* **309**, 167–195 (1985).
- Wellman, C. H. & Gray, J. The microfossil record of early land plants. *Phil. Trans. R. Soc. Lond. B* **355**, 717–732 (2000).
- Strother, P. K., Al-Hajri, S. & Traverse, A. New evidence for land plants from the lower Middle Ordovician of Saudi Arabia. *Geology* **24**, 55–58 (1996).
- Banks, H. P. The oldest vascular land plants: a note of caution. *Rev. Palaeobot. Palynol.* **20**, 13–25 (1975).
- Edwards, D. & Feehan, J. Records of *Cooksonia*-type sporangia from late Wenlock strata in Ireland. *Nature* **287**, 41–42 (1980).
- Edwards, D. & Wellman, C. H. in *Plants Invade the Land* (eds Gensel, P. G. & Edwards, D.) 3–28 (Columbia Univ. Press, New York, 2001).
- Droste, H. H. J. Stratigraphy of the Lower Palaeozoic Haima Supergroup of Oman. *GeoArabia* **2**, 419–468 (1997).
- Mohiuddin, U. *Quantitative Analysis of Ordovician Palynomorphs of Oman* Thesis (Univ. Sheffield).
- Edwards, D., Selden, P. A., Richardson, J. B. & Axe, L. Coprolites as evidence for plant–animal interaction in Siluro-Devonian terrestrial ecosystems. *Nature* **377**, 329–331 (1995).
- Taylor, W. A. Studies in cryptospore ultrastructure: variability in the tetrad genus *Tetrahedraletes* and type material of the dyad *Dyadospora murusattenuata*. *Rev. Palaeobot. Palynol.* **119**, 325–334 (2002).
- Pacini, E. in *Microspores: Evolution and Ontogeny* (eds Blackmore, S. & Knox, R. B.) 213–237 (Academic, London, 1990).
- Edwards, D. New insights into early land ecosystems: a glimpse of a Lilliputian world. *Rev. Palaeobot. Palynol.* **90**, 159–174 (1996).
- Richardson, J. B. in *Subsurface Palynostratigraphy of Northeast Libya* (eds El-Arnauti, B., Owens, B. & Thusu, B.) 89–109 (Garyounis Univ., Benghazi, Libya, 1988).
- Steenmans, P. Paléodiversification des spores et des cryptospores de l'Ordovician au Dévonien inférieur. *Geobios* **32**, 341–352 (1999).
- Taylor, W. A. Spores in earliest land plants. *Nature* **373**, 391–392 (1995).
- Edwards, D., Duckett, J. G. & Richardson, J. B. Hepatic characters in the earliest land plants. *Nature* **374**, 635–636 (1995).
- Wellman, C. H., Edwards, D. & Axe, L. Ultrastructure in laevigate hilate spores from sporangia and spore masses from the Upper Silurian and Lower Devonian of the Welsh Borderland. *Phil. Trans. R. Soc. Lond. B* **153**, 1983–2004 (1998).
- Wellman, C. H., Edwards, D. & Axe, L. Permanent dyads in sporangia and spore masses from the Lower Devonian of the Welsh Borderland. *Bot. J. Linn. Soc.* **127**, 117–147 (1998).
- Edwards, D., Wellman, C. H. & Axe, L. Tetrads in sporangia and spore masses from the Upper Silurian and Lower Devonian of the Welsh Borderland. *Bot. J. Linn. Soc.* **130**, 111–156 (1999).
- Graham, L. E. *Origin of Land Plants* (Wiley, New York, 1993).
- Brown, R. C. & Lemmon, B. E. in *Microspores: Evolution and Ontogeny* (eds Blackmore, S. & Knox, R. B.) 55–94 (Academic, London, 1990).
- Nickrent, D. L. *et al.* The relationships of vascular plants. *Phil. Trans. R. Soc. Lond. B* **355**, 847–855 (2000).
- Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants* (Smithsonian Institution, Washington, 1997).
- Nickrent, D. L. *et al.* Multigene phylogeny of land plants with special reference to the bryophytes and the earliest land plants. *Mol. Biol. Evol.* **17**, 1885–1895 (2000).
- Qiu, Y. L. & Lee, J. Transition to a land flora: a molecular phylogenetic perspective. *J. Phycol.* **36**, 799–802 (2000).
- Kroken, S. B., Graham, L. E. & Cook, M. E. Occurrence and evolutionary significance of resistant cell walls in charophytes and bryophytes. *Am. J. Bot.* **83**, 1241–1254 (1996).

Acknowledgements This research was supported by a NERC grant. We would like to thank Petroleum Development Oman and Oman Ministry of Oil and Gas for making the material available for study and for permission to publish.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to C.H.W. (c.wellman@sheffield.ac.uk).

Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins

Eric A. Gaucher¹, J. Michael Thomson^{2*}, Michelle F. Burgan³ & Steven A. Benner^{1,2,3}

¹NASA Astrobiology Institute, ²Department of Anatomy and Cell Biology, College of Medicine, and ³Department of Chemistry, University of Florida, Gainesville, Florida 32611-7200, USA

* Present address: Department of Cell and Developmental Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

Features of the physical environment surrounding an ancestral organism can be inferred by reconstructing sequences^{1–9} of ancient proteins made by those organisms, resurrecting these proteins in the laboratory, and measuring their properties. Here, we resurrect candidate sequences for elongation factors of the Tu family (EF-Tu) found at ancient nodes in the bacterial evolutionary tree, and measure their activities as a function of temperature. The ancient EF-Tu proteins have temperature optima of 55–65 °C. This value seems to be robust with respect

to uncertainties in the ancestral reconstruction. This suggests that the ancient bacteria that hosted these particular genes were thermophiles, and neither hyperthermophiles nor mesophiles. This conclusion can be compared and contrasted with inferences drawn from an analysis of the lengths of branches in trees joining proteins from contemporary bacteria¹⁰, the distribution of thermophily in derived bacterial lineages¹¹, the inferred G+C content of ancient ribosomal RNA¹², and the geological record combined with assumptions concerning molecular clocks¹³. The study illustrates the use of experimental palaeobiochemistry and assumptions about deep phylogenetic relationships between bacteria to explore the character of ancient life.

This year marks the 40th anniversary of the observation by Pauling and Zuckerkandl that it should be possible to infer the sequences of ancient proteins by comparing the sequences of their descendants¹⁴. Some 25 yr were required, however, before their vision of resurrecting ancient proteins for study was first realized^{1,2,4}. The properties of resurrected ancestral proteins have been used to correlate molecular behaviour with changing geology, ecology and physiology in mammals¹⁵, analyse the evolution of substrate specificity in biomedically important proteases⁵, and identify *in vitro* behaviours of proteins involved in inflammation and vision that are important to changing physiological function^{8,9}.

So far, however, experimental palaeobiochemistry has carried experimental scientists back in time only approximately 240 million years⁸. This has left untouched many of the most intriguing questions about ancient life. One of these relates to the role of thermophily in the history of life on Earth. Various models for environments in the Precambrian have suggested that the Earth was cold and covered with snow. Other models, inspired by the discovery of modern microorganisms that live at high temperatures, suggest that early bacteria may have been thermophiles, or possibly extreme thermophiles. Arguments based on indirect evidence, such as the lengths of branches of various trees, the G+C content of reconstructed ancestral rRNA, the possible cold temperature of early Earth, and the distribution of thermophily in contemporary taxa, have generated contradictory inferences.

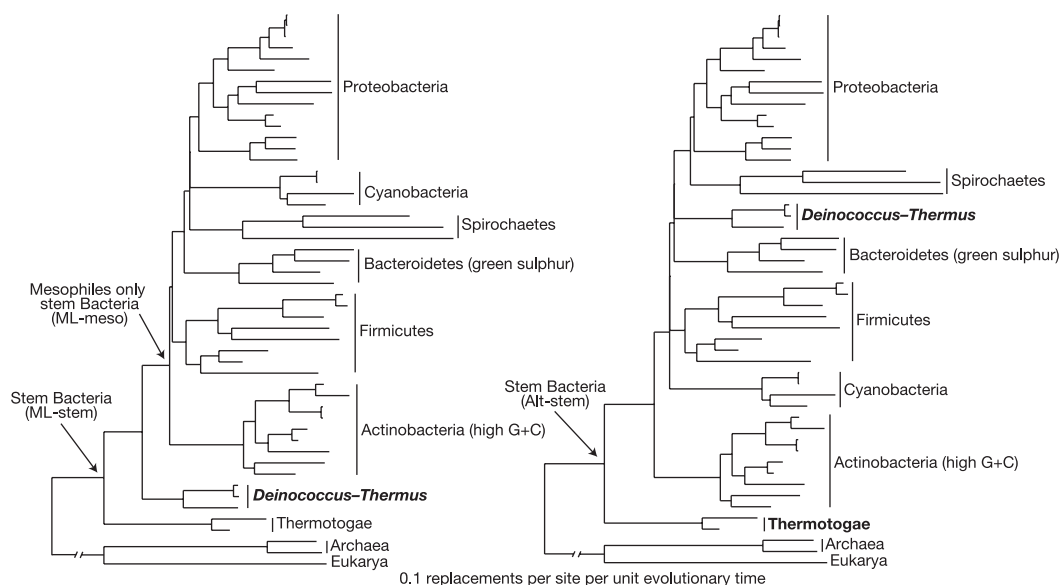


Figure 1 The two unrooted universal trees used to reconstruct ancestral bacterial sequences. Archaea and Eukarya serve to provide a node within the bacterial subtree from which ancient sequences can be inferred. Thermophilic lineages are highlighted in bold. Aquificaceae subfamily not shown. **a**, Maximum likelihood topology used to reconstruct

the stem elongation factors from bacteria (ML-stem), or most recent common ancestor of bacteria, and the ancestral sequence for mesophilic lineages only (ML-meso). **b**, Alternative topology used to reconstruct the stem elongation factors from bacteria (Alt-stem).

We reasoned that if we were able to reconstruct the sequences of ancestral proteins from bacteria that lived in the Precambrian, resurrect these proteins in the laboratory, and measure their thermal stabilities, we might be able to obtain direct evidence addressing the temperature(s) at which the particular ancestral bacteria lived. To be suited for the study, the protein would need to have temperature-dependent behaviour, with an optimum at physiological temperature. Furthermore, the rate of divergence of the protein must be slow enough, and the number of available derived and sibling protein sequences large enough, that the ancestral sequences can be reconstructed with minimal ambiguity. In addition, when the inevitable ambiguity is encountered, various sequences capturing that ambiguity must be sampled to see whether the interpretation, however made, is robust with respect to that ambiguity.

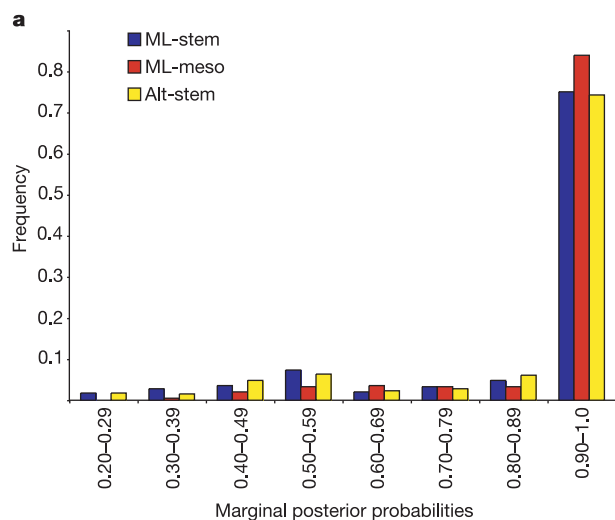
Elongation factor Tu (from Bacteria) and elongation factor 1A (from Archaea and Eukarya) are suitable proteins for such a study on all counts. EFs are G proteins that present charged aminoacyl-transfer RNAs to the ribosome during translation. Because of their relatively slow rates of sequence divergence, most character states of ancient EF sequences can be robustly reconstructed for proteins that lived on the order of a billion years ago. This has made them frequently useful for phylogenetic analysis. Furthermore, the optimal thermal stabilities of EFs correlate with the optimal growth temperature of the host organism. Thus, EFs from mesophiles, thermophiles and hyperthermophiles—defined as organisms that grow at 20–40, 40–80 and >80 °C, respectively, and represented by species of *Escherichia*, *Thermus* and *Thermotoga*—have temperature optima in their respective ranges^{16–18}. This is consistent with a previous study based on a large set of proteins in which a correlation coefficient of 0.91 was calculated between environmental temperatures of the host organisms and protein melting temperatures¹⁹.

To infer the sequences of EFs deep within the bacterial lineage, amino acid sequences of 50 EF-Tu proteins from various bacterial

lineages were collected, aligned and phylogenetically analysed. Because saturation at silent sites in the DNA sequence had occurred, amino acid sequences were used in the analysis. The differences in the rates of amino acid replacement at different sites in the sequence were captured using a gamma distribution²⁰. To support an analysis of the robustness of our interpretations with respect to plausible changes in our evolutionary models, two phylogenetic trees were used (see Methods). The first was constructed from EF-Tu sequences alone using a combination of phylogenetic tools (see Methods). The second was constructed from the literature, which contains various views of bacterial phylogeny (see, for example, ref. 11). These generally agree among themselves and with the EF-Tu tree. Where they differed, however, we extracted a tree that captured those differences (Fig. 1).

Candidate ancestral sequences were then reconstructed at nodes throughout the bacterial subtree. Marginal reconstructions, opposed to joint reconstructions, were calculated owing to our interest in comparing probabilities of multiple character states at a single interior node and selecting the character with the highest posterior probability²¹. The ‘most probabilistic ancestral sequence’ (MPAS) was then reconstructed by accepting at each site the amino acid with the highest posterior probability.

The MPASs for the two trees were found to be surprisingly similar to sequences from modern day *Aquifex*; only 4–6 amino acid replacements out of about 400 residues were inferred to have occurred from the most recent common ancestor of bacteria to the modern *Aquifex*. The placement of the branch leading to Aquificaceae at the base of the tree appeared, however, to be due



b

	Alt-stem	ML-meso	T.m.	T.a.	E.c.	G.s.
ML-stem	93	87	85	84	79	84
Alt-stem	-	89	84	80	80	83
ML-meso	-	-	74	78	82	84

Figure 2 Comparisons of reconstructed ancestral sequences. **a**, Distribution of the marginal posterior probabilities at sites for the three reconstructed ancestral nodes. The overall accuracies for sequence reconstructions of ML-stem, Alt-stem and ML-meso are predicted to be 88%, 88% and 92%, respectively²⁷. **b**, Examples of the per cent sequence identity between ancestral and modern proteins. T.m., *T. maritima*; T.a., *T. aquaticus*; E.c., *E. coli*; G.s., *Geobacillus stearothermophilus*.

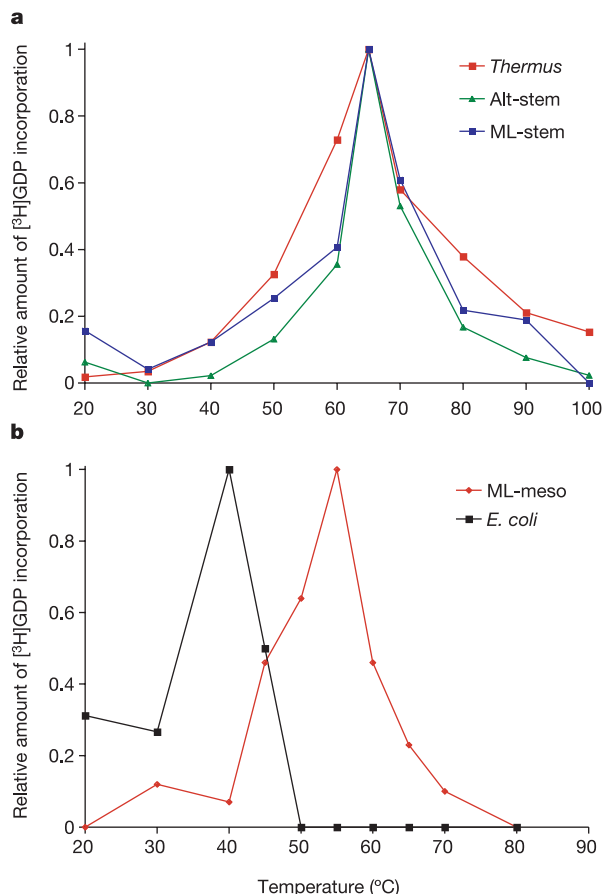


Figure 3 GDP-binding assay to test thermostability of ancestral and modern EF proteins. The amount of tritium-labelled GDP bound at 0 °C was subtracted from all other temperature values for a given protein. Shown is the relative amount of GDP bound compared with the amount bound at the optimal temperature for each protein.

to long-branch attraction, an artefact that has been much discussed^{13,22}. To test this, the 23 sites that displayed no variation within the Archaea/Eukarya subfamily, the Aquificaceae subfamily and the subfamily containing all other bacteria, but not conserved between the subfamilies, were removed from the analysis. The resulting analysis no longer places Aquificaceae at the base of the bacterial lineage. This is consistent with the working of long-branch attraction. To eliminate bias due to this artefact, ancestral sequences were recalculated from a data set that excluded *Aquifex* sequences.

Figure 1 shows the two topologies used to reconstruct ancestral sequences at the node representing the hypothetical organism laying near the presumed stem of the bacterial tree. The number of archaeal and eukaryal sequences (3–20) did not affect the amino acid reconstructions at these nodes: ML-stem (maximum likelihood stem for elongation factors in bacteria) and Alt-stem (alternative stem for the elongation factors in bacteria). The ancestral sequence at the node representing the most recent common ancestor of only mesophilic bacterial lineages was also reconstructed, and named ML-meso (maximum likelihood mesophiles only). This node captures a feature of models that have concluded that the last common ancestor of certain bacteria was mesophilic²². In all, these reconstructed ancestral sequences did not appear to be influenced by long-branch attraction or non-homogeneous modes of molecular evolution, such as changes in the mutability of individual sites in different branches of the bacterial subtree²⁰. Figure 2 presents the range of posterior probabilities for the reconstructed sequences and shows the sequence identity relating the putative sequences and their descendants. ML-stem and Alt-stem are most similar to the sequences of EFs from *Thermoanaerobacter tengcongensis* (a thermophile) and *Thermotoga maritima* (a hyperthermophile), respectively, and differ from each other by 28 amino acids. ML-meso is most similar to the sequence of EF from *Neisseria meningitidis* (a mesophile).

If we assume that similarity in sequence implies similarity in thermostability, we would predict that the stem bacterium was thermophilic or hyperthermophilic, and that the ancestral node constructed without considering thermophiles was mesophilic. To test these predictions based on this (unsubstantiated) assumption, genes encoding the ancestral sequences were synthesized, expressed in an *Escherichia coli* host and purified (see Methods). The thermostabilities of these ancestral EFs, and three representative EFs from contemporary organisms, were then assessed by measuring the ability of each to bind GDP across a range of temperatures.

Each resurrected protein behaved similarly (Fig. 3a). Both ML-stem and Alt-stem bound GDP with a temperature profile similar to that of the thermophilic EF from modern *Thermus aquaticus*, with optimal binding at about 65 °C. Although the sequence similarity was higher between Alt-stem and the modern hyperthermophilic *T. maritima*, the temperature profile of Alt-stem was not similar to that from *T. maritima*, which is maximally active up to at 85 °C (data not shown). The observation that the amino acid sequences of ML-stem and Alt-stem shared only 93% identity, but display the same thermostability profiles, suggests that inferences of this ancestral property are robust with respect to both varying topologies and ancestral character state predictions. This suggests, on the basis of these given evolutionary models, that the temperature in the palaeoenvironment of the ancient bacterium that hosted these reconstructed proteins was approximately 65 °C.

We then asked what inferences might be drawn from a resurrected EF whose sequence was reconstructed from the last common ancestral sequences of contemporary organisms that, for the most part, grow optimally at mesophilic temperatures. The temperature profile of the ancestral protein, which displayed a maximum at 55 °C, suggests that the ancestor of modern mesophiles lived at a higher temperature than its descendants (Fig. 3b). This result demonstrates that the behaviour of an ancestral protein need not be an average of the behaviours of its descendants, and suggests that

phylogenetic-based ancestral sequence reconstructions (per stirpes) should be preferred over consensus sequence reconstructions (per capita) and applied generally²³. The observation is important because it underscores the fact, well known in protein chemistry, that physical behaviour in a protein cannot be reliably predicted by a model that assumes amino acids at each site contribute to protein behaviour independently of the residues at all other sites. This, in turn, implies that an experiment in palaeobiochemistry can yield information beyond that yielded by analysis of descendant proteins alone.

Clearly, as more bacterial genomes are completed, we will need to exploit the resulting opportunity to extend these inferences to other nodes in the universal tree. Also, although our interpretation is robust with respect to the sampling of sequences from the extremes of the evolutionary models presented here, other plausible intermediate sequences will need to be resurrected and tested to establish this robustness more broadly. This notwithstanding, the results presented here show the value of experimental palaeobiochemical reconstructions to explore features in the lifestyles of ancient life forms, and will be particularly valuable when the relation of these ancient life forms is established relative to the root of the universal tree by other methods. □

Methods

Computational analyses

Completely sequenced genomes were used to determine the protein families that might support reconstructions in the bacterial tree most successfully. The MasterCatalog (EraGen Biosciences)¹⁵ was used as a starting point, as it presents an evolutionary model for all protein families in a recent version of GenBank. Gene trees were surveyed to identify the family that had the most sequences from the most diverse taxa with the least overall protein sequence divergence, using the PAM (point accepted mutations per 100 amino acids) distance metric. The family that represented the most potential for ancestral reconstruction was elongation factor Tu (Bacteria, EF-Tu) and 1A (Archaea/Eukarya, EF1A). This family was highly conserved, contained members from all of the complete genomes, and GenBank and Swiss-Prot databases contained diverse EF representatives. Sequences were aligned using ClustalW, and minor corrections were performed by hand. The list of organisms, GI and/or Swiss-Prot numbers, and multiple sequence alignment are available on request.

Two phylogenetic trees were used in the computation of ancestral amino acid character states for bacterial EF-Tu (about 400 residues). The maximum likelihood (ML) tree was generated as follows: a distance matrix was generated from the EF sequences using a gamma distribution to capture site-specific rate heterogeneity using the MEGA package²⁴. This distance matrix was then used to build trees inter-relating the sequences from individual lineages of bacteria using the minimum evolution criterion in the PAUP* package²⁵. The subtrees within the nine individual bacterial lineages were thus constrained, and the relationships between the lineages were then inferred using the maximum likelihood algorithm implemented within the MOLPHY package²⁶. The second tree was constructed by identifying in the literature plausible alternative bacterial phylogenies. Many of these combined analyses of multiple protein, DNA and rRNA sequences. Where these trees differed from the first tree, the differences were captured in an alternative tree (Alt).

The computation of ancestral amino acid states was performed using an empirical bayesian statistical framework that incorporated the gamma distribution (PAML)²⁷. For both the ML and alternative trees, ancestral character states were predicted using the Jones, Taylor and Thornton replacement matrix²⁸ (JTT) (See Supplementary Information). Using the Dayhoff²⁹ and WAG³⁰ (Whelan And Goldman) matrices had little effect on the predicted ancestral states.

Synthetic gene construction and protein expression

The ancestral genes were synthesized piecemeal by polymerase chain reaction using complementary 50-nucleotide oligonucleotides (MWG Biotech) with 15–20 base-pair overlap. Ancestral and extant (*E. coli*, *T. aquaticus* and *T. maritima*) EF genes were cloned in Topo-TA (Invitrogen). Errors resulting from primer synthesis or PCR were fixed using standard site-directed mutagenesis. All genes were sequenced with two times coverage. The genes were subsequently cloned and expressed in the TYB11 vector (IMPACT System; intein-mediated purification with an affinity chitin-binding tag), and purified according to the manufacturer (New England Biolabs). The proteins were eluted from the chitin affinity column in a buffer comprising 20 mM Tris-HCl (pH 8.5), 500 mM NaCl, 10 mM MgCl₂, 5 μM GDP and 1 μM phenylmethylsulphonyl fluoride and stored at –20 °C. The samples were filtered and concentrated using Centricon YM-30 (Amicon). SDS-PAGE verified the isolation of a single band of appropriate size: about 44 kDa.

GDP-binding assay

The thermostability of ancestral and extant proteins was determined based on an ability to bind nucleotide across a range of temperatures^{16–18}. Aliquots (40 μl) of a solution

containing 1 μ M EF protein in 20 mM Tris-HCl (pH 8.5), 500 mM NaCl, 10 mM MgCl₂ and 2.5 μ M [³H]GDP (specific activity approximately 11.5 Ci mmol⁻¹) (Amersham Pharmacia Biotech) were assayed at different temperatures, first at 10 °C intervals between 0 and 100 °C and then $\pm$ 5 °C on either side of the temperature optimum. For the ML-meso putative ancestor, binding at 5 °C intervals between 40 and 70 °C was also determined. The amount of [³H]GDP bound to EF was determined as previously described¹⁶.

Received 7 April; accepted 30 July 2003; doi:10.1038/nature01977.

- Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. & Wilson, A. C. Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing. *Nature* **345**, 86–89 (1990).
- Stackhouse, J., Presnell, S. R., McGeehan, G. M., Nambiar, K. P. & Benner, S. A. The ribonuclease from an extinct bovid. *FEBS Lett.* **262**, 104–106 (1990).
- Adey, N. B., Tollefsbol, T. O., Sparks, A. B., Edgell, M. H. & Hutchison, C. A. Molecular resurrection of an extinct ancestral promoter for mouse L1. *Proc. Natl Acad. Sci. USA* **91**, 1569–1573 (1994).
- Jermann, T. M., Opitz, J. G., Stackhouse, J. & Benner, S. A. Reconstructing the evolutionary history of the aridactyl ribonuclease superfamily. *Nature* **374**, 57–59 (1995).
- Chandrasekharan, U. M., Sanker, S., Glynnias, M. J., Karnik, S. S. & Husain, A. Angiotensin II forming activity in a reconstructed ancestral chymase. *Science* **271**, 502–505 (1996).
- Golding, G. B. & Dean, A. M. The structural basis of molecular adaptation. *Mol. Biol. Evol.* **15**, 355–369 (1998).
- Miyazaki, J. et al. Ancestral residues stabilizing 3-isopropylmalate dehydrogenase of an extreme thermophile: Experimental evidence supporting the thermophilic common ancestor hypothesis. *J. Biochem.* **129**, 777–782 (2001).
- Chang, B. S., Jonsson, K., Kazmi, M. A., Donoghue, M. J. & Sakmar, T. P. Recreating a functional ancestral archosaur visual pigment. *Mol. Biol. Evol.* **19**, 1483–1489 (2002).
- Zhang, J. Z. & Rosenberg, H. F. Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. *Proc. Natl Acad. Sci. USA* **99**, 5486–5491 (2002).
- Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
- Hugenholtz, P., Goebel, B. M. & Pace, N. R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**, 4765–4774 (1998).
- Galtier, N., Tourasse, N. & Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**, 220–221 (1999).
- Cavalier-Smith, T. The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* **52**, 7–76 (2002).
- Pauling, L. & Zuckerkandl, E. Chemical paleogenetics: molecular “restoration studies” of extinct forms of life. *Acta Chem. Scand. A* **17**, S9–S16 (1963).
- Benner, S. A., Caraco, M. D., Thomson, J. M. & Gaucher, E. A. Planetary biology—paleontological, geological, and molecular histories of life. *Science* **296**, 864–868 (2002).
- Arai, K.-I., Kawakita, M. & Kaziro, Y. Studies on polypeptide elongation factors from *Escherichia coli*. *J. Biol. Chem.* **247**, 7029–7037 (1972).
- Nock, S. et al. Properties of isolated domains of the elongation factor Tu from *Thermus thermophilus* HB8. *Eur. J. Biochem.* **234**, 132–139 (1995).
- Sanangelantoni, A. M., Cammarano, R. & Tiboni, O. Manipulation of the tuf gene provides clues to the localization of sequence element(s) involved in the thermal stability of *Thermotoga maritima* elongation factor Tu. *Microbiol.* **142**, 2525–2532 (1996).
- Gromiha, M. M., Oobatake, M. & Sarai, A. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys. Chem.* **82**, 51–67 (1999).
- Gaucher, E. A., Miyamoto, M. M. & Benner, S. A. Function-structure analysis of proteins using covarian-based evolutionary approaches: Elongation factors. *Proc. Natl Acad. Sci. USA* **98**, 548–552 (2001).
- Yang, Z. H., Kumar, S. & Nei, M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641–1650 (1995).
- Brochier, C. & Philippe, H. Phylogeny: a non-hyperthermophilic ancestor for bacteria. *Nature* **417**, 244 (2002).
- Gaschen, B. et al. Diversity considerations in HIV-1 vaccine selection. *Science* **296**, 2354–2360 (2002).
- Kumar, S., Tamura, K. & Nei, M. MEGA—Molecular evolutionary genetics analysis software for microcomputers. *Comput. Appl. Biosci.* **10**, 189–191 (1994).
- Swofford, D. L. PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)* Version 4.10 (Sinauer Associates, Sunderland, Massachusetts, 2002).
- Adachi, J. & Hasegawa, M. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**, 1–150 (1996).
- Yang, Z. PAML: a program package for phylogenetic analyses by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1999).
- Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequence. *Comput. Appl. Biosci.* **8**, 275–282 (1992).
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. *Atlas of Protein Sequence and Structure* 345–358 (National Biomedical Research Foundation, Washington DC, 1978).
- Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Miyamoto, J. Aris, A. Falcon, S. Sassi, C. West and Z. Yang for discussions and assistance with our research. We also thank C. Knudson and A. M. Sanangelantoni for providing EF clones. Funding is provided by the National Research Council and NASA's Astrobiology Institute (E.A.G.), and the NIH and NASA (S.A.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.A.G. (gaucher@ufl.edu). Inferred ancestral sequences have been deposited in GenBank under accession numbers AY305395–AY305397.

Patterns of predation in a diverse predator–prey system

A. R. E. Sinclair¹, Simon Mduma^{1,2} & Justin S. Brashares¹

¹Centre for Biodiversity Research, 6270 University Boulevard, University of British Columbia, Vancouver, V6T 1Z4, Canada

²Tanzania Wildlife Research Institute, Box 661, Arusha, Tanzania

There are many cases where animal populations are affected by predators and resources in terrestrial ecosystems^{1–3}, but the factors that determine when one or the other predominates remain poorly understood^{4–5}. Here we show, using 40 years of data from the highly diverse mammal community of the Serengeti ecosystem, East Africa, that the primary cause of mortality for adults of a particular species is determined by two factors—the species diversity of both the predators and prey and the body size of that prey species relative to other prey and predators. Small ungulates in Serengeti are exposed to more predators, owing to opportunistic predation, than are larger ungulates; they also suffer greater predation rates, and experience strong predation pressure. A threshold occurs at prey body sizes of ~150 kg, above which ungulate species have few natural predators and exhibit food limitation. Thus, biodiversity allows both predation (top-down) and resource limitation (bottom-up) to act simultaneously to affect herbivore populations. This result may apply generally in systems where there is a diversity of predators and prey.

The influence of predation and resource availability on population dynamics has long been a focus of ecological research⁶. Yet, we know little about how these top-down and bottom-up forces work together to structure diverse ecosystems^{4–6,7}. This is particularly true for mammal communities, where effects of predation and resource limitation are usually investigated on single species or in simple systems with low species diversity^{8,9}. We examined patterns of predation in the diverse Serengeti ecosystem using data on ungulate populations before, during and after a period of predator removal.

The Serengeti ecosystem (34–36° E, 1–4° S) in Tanzania and Kenya, East Africa, is composed of open grassland and savannah. It supports 28 species of ungulates and 10 species of large carnivores that prey on them¹⁰. In any one habitat, there can be seven co-

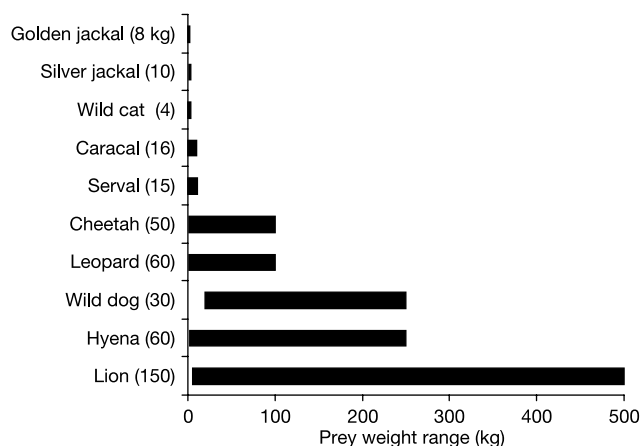


Figure 1 The range of weights of mammal prey consumed by carnivores of different sizes in the Serengeti ecosystem. There is a large overlap in diet at small prey sizes. Data are from our unpublished observations and published sources^{17,20}.

occurring species of canid and felid carnivores feeding on an array of ungulates. The carnivores show a characteristic pattern in the sizes of their typical prey. Each predator species tends to differ from others in preferring a narrow range of prey sizes within a broader diet range (Supplementary Information). Thus, the lion's (*Panthera leo*) preferred prey is wildebeest (*Connochaetes taurinus*) and zebra (*Equus quagga*) (170–250 kg; 38% of animals consumed), but 44% of the diet is made up of smaller prey (2–45 kg), while buffalo (*Syncerus caffer*) and giraffe (*Giraffa camelopardalis*) comprise 5–15% (refs 11, 12; approximate percentages in terms of biomass consumed are 63, 14 and 15–24, respectively). Each carnivore species includes prey outside their preferred size range but is inefficient at catching such prey. So, while lions are less efficient at catching gazelle (*Gazella sp.*) than catching wildebeest¹³, they do so if opportunities present themselves. Hence, gazelle are less vulnerable to lion predation, despite their small size, than are wildebeest¹³. Overall, the diet niche of smaller carnivores is nested within that of larger carnivore species (Fig. 1).

Smaller-bodied ungulates suffer predation from many more predators than do larger ungulates (Fig. 2). Small species of antelope, such as oribi (*Ourebia ourebi*), are prey to five species of cats, two canids, hyena (*Crocuta crocuta*) and many smaller carnivores in Serengeti. We observed this range of predators with oribi where predation was caused by lion (14%), hyena (15%), leopard (*Panthera pardus*; 43%), silver-backed jackal (*Canis mesomelas*; 9%), baboon (*Papio anubis*; 7%), eagle (7%) and python (6%)¹⁴. Medium-sized antelopes (for example, wildebeest) are prey to three felids, the wild dog (*Lycaon pictus*) and hyena, while larger ungulates such as buffalo and giraffe are depredated by lion only¹⁵. Very large herbivores (that is, elephant (*Loxodonta africana*), rhinoceros (*Diceros bicornis*) and hippopotamus (*Hippopotamus amphibius*)) almost never suffer predation as adults and only rarely as juveniles.

Approximate densities per km² of the major predators in northern Serengeti are lion 0.22, hyena 0.5, leopard 0.12, cheetah (*Acinonyx jubatus*) 0.08, wild dog 0.02, and silver-backed jackal 0.16, providing a total of 1.1 predators per km² (ref. 16). These differences in density among species combined with differences in predation rates suggest that some carnivores have a greater effect on prey populations than others. Actual predator densities are likely to be much higher if the golden jackal (*Canis aureus*) and other smaller carnivores (1–10 kg) are included, but these data are not yet complete. Preliminary counts suggest these smaller predators occur at densities at least 2–3 times higher than those of the largest carnivores. Small prey such as oribi or Thomson's gazelle (*Gazella rufifrons*) at approximate densities of 3 per km² (see below) are faced

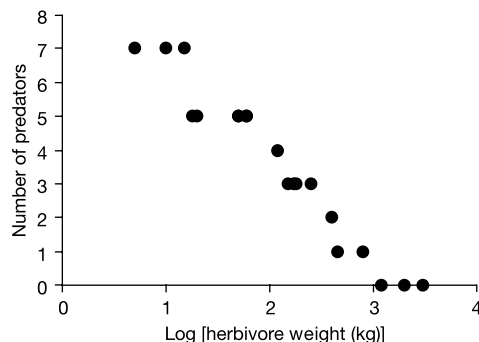


Figure 2 The number of mammal carnivore species that prey upon the savannah ungulates of different body sizes in Serengeti. Adult female body sizes are from published sources²⁶ and provided in Supplementary Information.

with a predator:prey ratio of at least 1:3, which is extremely high. Thus, smaller ungulates should experience greater mortality from opportunistic predation, and we predict their populations should be severely affected by predation.

This prediction was tested with data from long-term studies of ungulate mortality rates and their causes. Using the long-term recruitment and mortality rates of ungulate populations and direct measures of mortality from telemetry^{17–20}, and information on predator diets^{11,12,21}, we calculated the proportion of adult mortality of ungulates accounted for by predation (Fig. 3). All records of adult mortality of impala and oribi indicate predation. In contrast, only about 5% of giraffe and 23% of buffalo adult mortality is caused by predators, and the proportion due to predation declines to zero for the largest ungulates. These values reflect adult mortality only, but predation at the juvenile stage will have an additional effect on most herbivore populations, particularly for smaller-bodied species. The relationship between ungulate body size and predation rate is not linear. Above a threshold body size of about 150 kg the proportion of annual adult mortality caused by predators declines rapidly as body size increases, indicating a sharp transition from top-down to bottom-up regulatory processes.

The pattern of predation mortality relative to prey size (Fig. 3) is predicted from differing levels of opportunistic predation through exposure to different numbers of predator species (Fig. 2). Such a pattern may also be predicted if small-bodied prey are uniformly more vulnerable to all predator species. However, data on predation rates, and predator diet and efficiencies^{11–13,15,16,21,22}, show that prey are differentially vulnerable to different predators, making this alternative mechanism less probable.

A natural experiment allowed a further test of this relationship between herbivore size and mode of population regulation. In a study area of northern Serengeti that has been monitored since 1966, poaching and indiscriminate poisoning removed the majority of carnivores (notably lion, hyena and jackals from evidence on observed carcasses and absence of calls) for an eight-year period (1980–87)¹⁹. In the contiguous Mara Reserve of Kenya, a part of the same ecosystem, predator populations remained intact²³. During the predator removal period, five smaller-bodied species (oribi, Thomson's gazelle, warthog (*Phacochoerus africanus*), impala (*Aepyceros melampus*) and topi (*Damaliscus lunatus*)) for which data were available, increased markedly in density in the predator removal area relative to their populations in the non-removal area

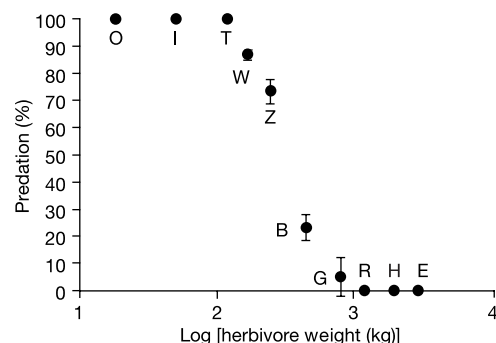


Figure 3 The proportion of annual adult mortality accounted for by predation in ten non-migratory ungulate populations for which data were available in the Serengeti ecosystem. There is a threshold in body size of about 150 kg above which predator limitation switches to food limitation. Error bars are 95% confidence limits. Species are: O, oribi; I, impala; T, topi; W, wildebeest; Z, zebra; B, African buffalo; G, giraffe; R, black rhino; H, hippo; E, African elephant. Data are provided in Supplementary Information.

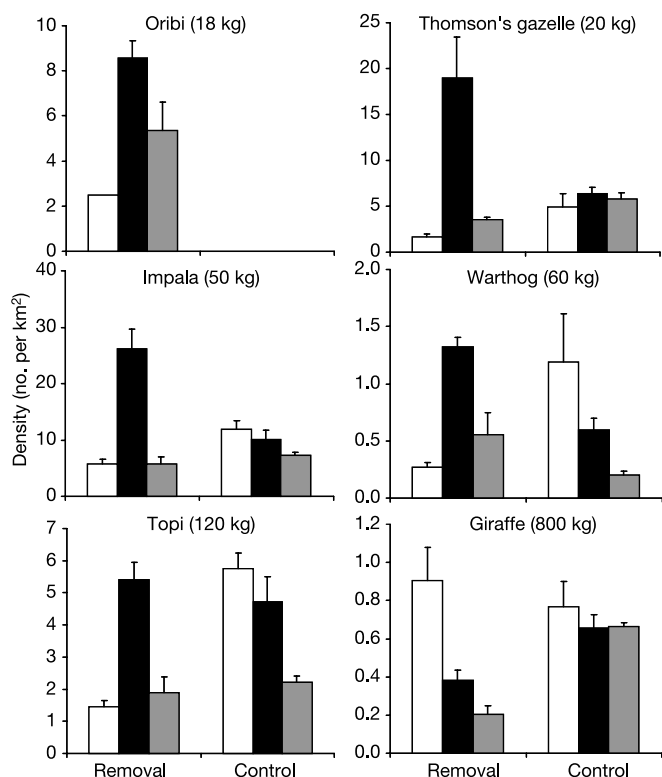


Figure 4 Densities of six ungulates in a predator removal area in northern Serengeti^{17,27} compared with an adjacent control in the Mara Reserve²³. Period before removal (open bar) covered 1967–80, the period of low predator numbers due to poaching (solid bar) covered 1981–87, and the period after predators returned (shaded bar) included years from 1989 onwards¹⁷. The five smaller species (<150 kg) increased during the predator removal period relative to the control, whereas the larger giraffe did not respond by increasing. These results are predicted from the pattern in Fig. 3. All comparisons of before versus during, and during versus after, periods differed significantly in the removal area ($P = 0.05$ – 0.001 calculated from standard census formulae^{28,29} or t -tests assuming unequal variance). All comparisons in the control area were not significant except for warthog and topi ($P < 0.05$) that declined owing to other system-wide processes. There is no control for oribi because they do not occur in the Mara Reserve. Census locations and methods are described elsewhere^{17,23,30}. Error bars are 1 standard error. Data are provided in Supplementary Information.

(Fig. 4). These populations declined once predators returned to the area. In contrast, abundance of one of the largest ungulates, the giraffe, did not increase in the predator removal area over the eight-year period (Fig. 4). These results conform to the prediction from Fig. 3 that species below a threshold body size (in this case <150 kg) should be predator limited whereas giraffe should not. If the loss of one predator species were to be compensated by increased predation from another, then we would not have observed the increase in prey populations, especially those of very small prey (Fig. 4).

In summary, evidence from predator diets, predation efficiencies and mortality of prey indicate that the preferences of carnivores impose the pattern of predation pressure on prey populations. This pattern is related to two features of scale—the nested diet niches of carnivores, and the body size of the prey. These two scaling factors result in heavy predation pressure (top-down processes) for small prey and resource limitation (bottom-up processes) for large prey in this diverse prey community. There exists a threshold in body size beyond which the cause of mortality switches to bottom-up control

in larger species. These results suggest a mechanism by which the diversity or abundance of predators and the diversity of prey may determine the effect of predation on populations. We hypothesize that the loss of such diversity could disrupt the interaction of these processes and result in outbreaks and extinctions^{2,7,24,25}. We suggest that these patterns of predation will emerge in other terrestrial systems that display similar complexity and scaling of predators and prey. □

Received 9 June; accepted 18 July 2003; doi:10.1038/nature01934.

- Persson, L. Trophic cascades: Abiding heterogeneity and the trophic level concept at the end of the road. *Oikos* **85**, 385–397 (1999).
- Terborgh, J. *et al.* Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1926 (2001).
- Sinclair, A. R. E. *et al.* Testing hypotheses of trophic level interactions: A boreal forest ecosystem. *Oikos* **89**, 313–328 (2000).
- Polis, G. A. Why are parts of the world green: Multiple factors control productivity and the distribution of biomass. *Oikos* **86**, 3–15 (1999).
- Dyer, L. A. & Letourneau, D. K. Trophic cascades in a complex terrestrial community. *Proc. Natl Acad. Sci. USA* **96**, 5072–5076 (1999).
- Hairston, N. G., Smith, F. E. & Slobodkin, L. B. Community structure, population control, and competition. *Am. Nat.* **94**, 421–424 (1960).
- Hunter, M. D. & Price, P. W. Playing chutes and ladders: Heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. *Ecology* **73**, 724–732 (1992).
- Post, E., Peterson, R. O., Stenseth, N. C. & McLaren, B. E. Ecosystem consequences of wolf behavioural response to climate. *Nature* **401**, 905–907 (1999).
- Schoener, T. W. & Spiller, D. A. Devastation of prey diversity by experimentally introduced predators in the field. *Nature* **381**, 691–694 (1996).
- Sinclair, A. R. E. & Norton-Griffiths, M. (eds) *Serengeti: Dynamics of an Ecosystem* (Univ. Chicago Press, Chicago, 1979).
- Shaller, G. B. *The Serengeti Lion: A Study of Predator-Prey Relations* (Univ. Chicago Press, Chicago, 1972).
- Bertram, B. C. R. in *Serengeti: Dynamics of an Ecosystem* (eds Sinclair, A. R. E. & Norton-Griffiths, M.) 221–248 (Univ. Chicago Press, Chicago, 1979).
- Elliot, J. P., Cowan, I. M. T. & Holling, C. S. Prey capture of the African lion. *Can. J. Zool.* **55**, 1811–1828 (1977).
- Brashares, J. S. & Arcese, P. in *The Mammals of Africa* (eds Kingdon, J., Happold, D. & Butynski, T.) (Academic, London, in the press).
- Scheel, D. & Packer, C. in *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 299–314 (Univ. Chicago Press, Chicago, 1995).
- Hofer, H. & East, M. in *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 332–363 (Univ. Chicago Press, Chicago, 1995).
- Sinclair, A. R. E. in *Serengeti: Dynamics of an Ecosystem* (eds Sinclair, A. R. E. & Norton-Griffiths, M.) 82–103 (Univ. Chicago Press, Chicago, 1979).
- Sinclair, A. R. E. Does interspecific competition or predation shape the African ungulate community? *J. Anim. Ecol.* **54**, 899–918 (1985).
- Sinclair, A. R. E. in *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 194–219 (Univ. Chicago Press, Chicago, 1995).
- Mduma, S. A. R., Sinclair, A. R. E. & Hilborn, R. Food regulates the Serengeti wildebeest: A 40-year record. *J. Anim. Ecol.* **68**, 1101–1122 (1999).
- Kruuk, H. *The Spotted Hyena: A Study of Predation and Social Behavior* (Univ. Chicago Press, Chicago, 1972).
- Lamprecht, J. On diet, foraging behavior and interspecific food competition of jackals in the Serengeti National Park, East Africa. *Z. Säugetierk.* **43**, 210–223 (1978).
- Broten, M. D. & Said, M. in *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 169–193 (Univ. Chicago Press, Chicago, 1995).
- Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecol. Lett.* **5**, 558–567 (2002).
- Dyer, L. A. & Letourneau, D. K. Top-down and bottom-up diversity cascades in detrital vs. living food webs. *Ecol. Lett.* **6**, 60–68 (2003).
- Nowak, R. M. *Walker's Mammals of the World* (Johns Hopkins, Baltimore, 1999).
- Mduma, S. A. R. in *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 220–230 (Univ. Chicago Press, Chicago, 1995).
- Jolly, G. M. Sampling methods for aerial census of wildlife populations. *E. Afr. Agric. For. J.* **34**, 46–49 (1969).
- Cochran, W. G. *Sampling Techniques* (Wiley, New York, 1963).
- Campbell, K. & Borner, M. in *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 117–145 (Univ. Chicago Press, Chicago, 1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Tanzania Wildlife Research Institute and Tanzania National Parks for facilitating our research in Serengeti. We thank NSERC Canada and the National Geographic Society for funding.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.R.E.S. (sinclair@zoology.ubc.ca).

The evolutionary inheritance of elemental stoichiometry in marine phytoplankton

Antonietta Quigg¹, Zoe V. Finkel¹, Andrew J. Irwin¹, Yair Rosenthal^{1,2}, Tung-Yuan Ho⁵, John R. Reinfelder³, Oscar Schofield^{1,4}, Francois M. M. Morel⁵ & Paul G. Falkowski^{1,2}

¹Environmental Biophysics and Molecular Ecology Program, Institute of Marine and Coastal Sciences, ²Department of Geological Sciences, ³Department of Environmental Sciences, and ⁴Coastal Ocean Observation Laboratory, Rutgers University, New Jersey 08901, USA

⁵Department of Geosciences, Princeton University, New Jersey 08544, USA

Phytoplankton is a nineteenth century ecological construct for a biologically diverse group of pelagic photoautotrophs that share common metabolic functions but not evolutionary histories¹. In contrast to terrestrial plants, a major schism occurred in the evolution of the eukaryotic phytoplankton that gave rise to two major plastid superfamilies^{2–4}. The green superfamily appropriated chlorophyll *b*, whereas the red superfamily uses chlorophyll *c* as an accessory photosynthetic pigment⁵. Fossil evidence suggests that the green superfamily dominated Palaeozoic oceans. However, after the end-Permian extinction, members of the red superfamily rose to ecological prominence. The processes responsible for this shift are obscure. Here we present an analysis of major nutrients and trace elements in 15 species of marine phytoplankton from the two superfamilies. Our results indicate that there are systematic phylogenetic differences in the two plastid types where macronutrient (carbon:nitrogen:phosphorus) stoichiometries primarily reflect ancestral pre-symbiotic host cell phenotypes, but trace element composition reflects differences in the acquired plastids. The compositional differences between the two plastid superfamilies suggest that changes in ocean redox state strongly influenced the evolution and selection of eukaryotic phytoplankton since the Proterozoic era.

To examine patterns of variation in biochemical composition, 15 species of marine eukaryotic phytoplankton from five phyla were grown using axenic techniques under identical conditions ($19 \pm 1^\circ\text{C}$; 12 h light/dark cycle; $250 \mu\text{mol photons m}^{-2} \text{s}^{-1}$). The average bulk elemental composition of $\text{C}_{125}\text{N}_{16}\text{P}_1$, by atoms, for the ensemble of phytoplankton (see Supplementary Fig. 1) was remarkably similar to the canonical stoichiometry ($\text{C}_{106}\text{N}_{16}\text{P}_1$) first recognized by Redfield⁶ as early as 1934 to have a causal correlation with the NO_3^- and PO_4^{3-} ratio in the ocean interior. The basic concept of the Redfield ratio became a cornerstone of ocean models in the ensuing decades; it represents an emergent property that integrates biogeochemical cycles on basin scales over tens of thousands of years⁷. It was clearly understood however, that “The ratios hold very well for the nitrate and phosphate in ocean waters, but, since they represent the net effect of biological activity, marked deviations from the ratios may be found in individual types of organisms”⁸.

Indeed, there are systematic variations in C:N:P ratios between different eukaryotic phyla and superfamilies. The differences in C:P and C:N ratios between phyla and superfamilies (Fig. 1) are statistically significant ($P < 0.01$; see Supplementary Table 1). Whereas N:P ratios differentiate between the superfamilies ($P < 0.01$), this was not the case for phyla (Fig. 1). Experimental evidence suggests that differences in cellular N:P composition between marine eukaryotic phytoplankton is a reflection of the ratio of total cell protein to ribosomal RNA^{9–11}, but not genome size. That we can differentiate between phyla and superfamilies supports

the suggestion that the C:N:P composition of particulate organic matter is dependent on the species composition of a phytoplankton assemblage. Hence, given the species composition of a phytoplankton assemblage, one could predict with reasonable precision the average C:N:P ratio of the ensemble community, but not vice versa.

By extending the compositional analysis to include the trace elements Fe, Mn, Zn, Cu, Cd, Co and Mo (see Supplementary Fig. 1), we were able to further differentiate between eukaryotic marine phytoplankton phyla (see Supplementary Table 1) and superfamilies ($P < 0.001$; analysis of variance (ANOVA)). To define stoichiometric elemental profiles, we considered the trace elemental data in aggregate using principal component analysis (PCA). The first two principal components clearly separate the superfamilies by plastid type, and often by phyla (Fig. 2); 72% of the total variance is accounted for by the PCA. Most of the variability within the green superfamily is orthogonal to that in the red superfamily, suggesting that trace elemental profiles reflect plastid origins. Trace elemental concentrations normalized to either carbon or cell volume returned statistically similar results. A linear discriminant model¹² was used to classify species as either a member of the red or green plastid superfamily. Purely on the basis of elemental composition, we were

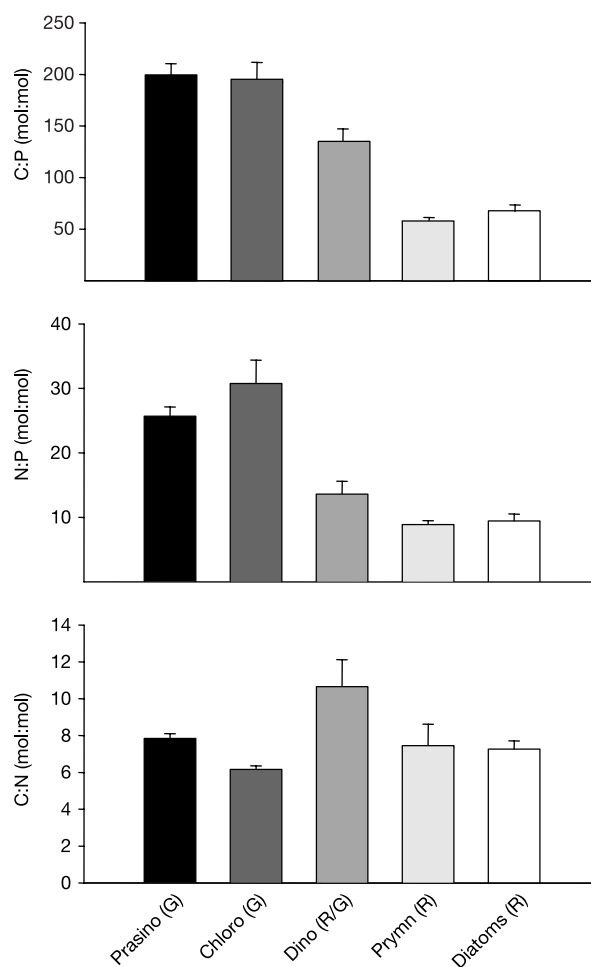


Figure 1 C:N:P composition varies between phyla and superfamilies. Phytoplankton C:P, N:P and C:N (mol:mol) ratios are grouped phylogenetically—Prasinophyceae (Prasino) and Chlorophyceae (Chloro) are members of the green (G) plastid superfamily whereas Dinophyceae (Dino), Prymnesiophyceae (Prymn) and Bacillariophyceae (Diatoms) are members of the red (R) plastid superfamily. Error bars indicate standard errors.

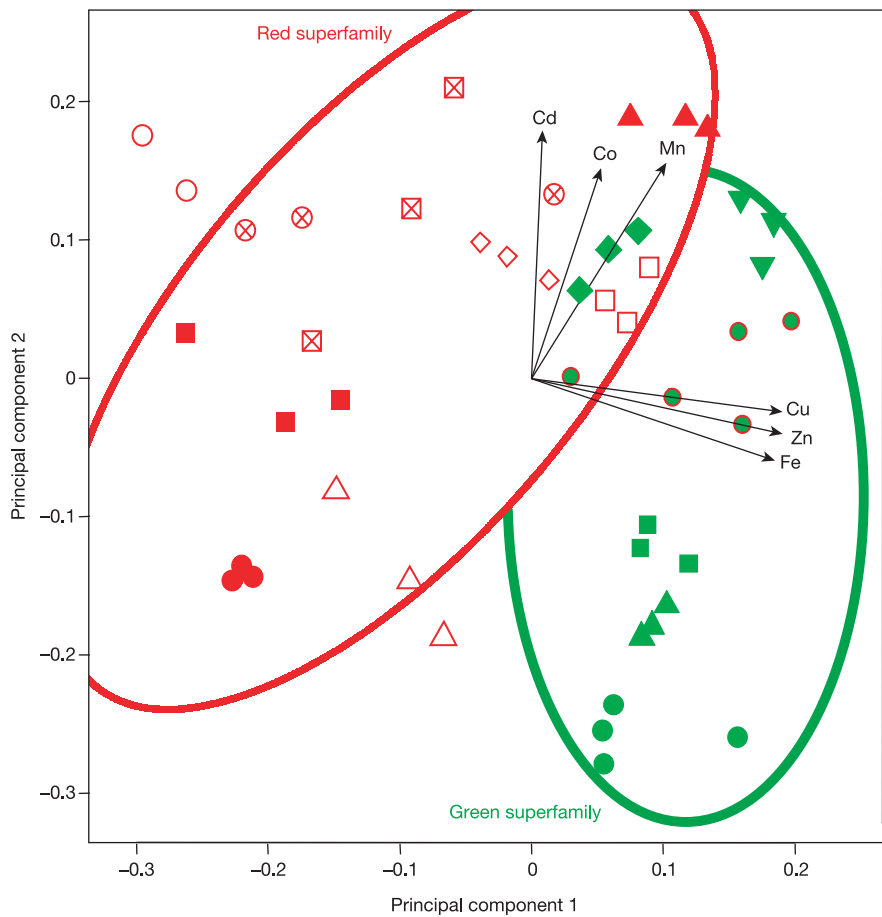


Figure 2 Elemental profiles differentiate the green and red superfamilies. Principal component analysis of all 15 species of phytoplankton examined clusters them according to their respective superfamilies, and to a lesser degree, phyla. The principal component axes are weighted sums of the log element:P ratios (log mmol:mol P) which are shown on the figure abbreviated by the element's symbol only. The first (PC1) and second (PC2) principal components— $PC1 = 0.53 \text{ Fe} + 0.29 \text{ Mn} + 0.56 \text{ Zn} + 0.55 \text{ Cu} + 0.15$

$Co + 0.025 \text{ Cd}$; $PC2 = -0.20 \text{ Fe} + 0.53 \text{ Mn} - 0.14 \text{ Zn} - 0.12 \text{ Cu} + 0.52 \text{ Co} + 0.61 \text{ Cd}$ —account for 42% and 30% of the variance respectively. Species symbols (Table 1) are centred on their location; the ellipses are used as an additional visual guide to the clustering. The length of each vector corresponds to its weighting in the first two principal components.

able to make the correct classification 85% of the time (averaged over 100 tests). Differences in the trace elemental composition between superfamilies are the consequence of environmental selection pressures on components of the photosynthetic apparatus where metal

substitutions are possible. For example, at some point, members of the green superfamily lost their ability to express the Fe metallo-protein cytochrome c_6 , and substituted plastocyanin (Cu) to ferry electrons from cytochrome b_6/f to photosystem I (ref 13). Despite this metal substitution, the Fe quotas in the green superfamily were

Table 1 Phytoplankton phyla examined and appearance in geological record^a

Geological era	Phyla	Species/clones
Proterozoic (2,500–544 Myr)	Prasinophyceae (G) (1,200 Myr)	<i>Pyramimonas parkeae</i> (CCMP 724; inverted green triangle) <i>Tetraselmis</i> sp. (CCMP 1639; green diamond) <i>Pycnococcus provasolii</i> (CCMP 1203; green square)
	Chlorophyceae (G) (1,000 Myr)	<i>Dunaliella tertiolecta</i> (CCMP 1320; green triangle) <i>Nannochloris atomus</i> (CCMP 509; green circle)
	Dinophyceae (R) (440 Myr)	<i>Gymnodinium chlorophorum</i> (DIN 3; red and green circle)* <i>Thoracosphaera heimii</i> (CCMP 1375; filled red square) <i>Prorocentrum minimum</i> (CCMP 1329; filled red triangle) <i>Amphidinium carterae</i> (CCMP 1314; filled red circle)
Palaeozoic (544–245 Myr)		
Mesozoic 245–65 Myr	Prymnesiophyceae (R) (210 Myr)	<i>Emiliana huxleyi</i> (ASM1; red circle and cross) <i>Gephyrocapsa oceanica</i> (JS1A; red square and cross)
	Bacillariophyceae (R) (120 Myr)	<i>Ditylum brightwellii</i> (CCMP 358; open red square) <i>Thalassiosira eccentrica</i> (CCMP 1800; open red square) <i>Thalassiosira weissflogii</i> (CCMP 1336; open red diamond) <i>Nitzschia brevirostris</i> (CCMP 551; open red triangle)

(G) and (R) indicate phyla associated with green and red plastid superfamilies respectively. Clones were purchased from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP). *Gymnodinium chlorophorum* was obtained from the Algalbank Collection (France). *Emiliana huxleyi* and *G. oceanica* were provided by I. Probert (Universite Caen Basse Normandie). Symbol descriptions in the third column are of those used in Fig. 2. Myr, million years ago. **Gymnodinium chlorophorum* has a chlorophyll *b*-containing (green) plastid rather than chlorophyll *c*-containing (red) plastid.

consistently higher than those for the red superfamily (Fig. 2), possibly due to higher ratios of photosystem I (containing 12 atoms of Fe) to II (containing 3 atoms of Fe) in the former¹⁴. Phytoplankton in the red superfamily typically had higher Cd and Co to P ratios (Fig. 2). The restoration of growth by Cd or Co in several species of Zn-limited diatoms and coccolithophorids may involve substitution of these elements for Zn in carbonic anhydrase^{15,16} and perhaps other proteins. Hence, species-specific elemental profiles reflect differences in the biochemistry of the plastids.

The separation of the superfamilies in the PCA (Fig. 2) according to their plastid type does not seem to be a direct consequence of the information encoded in extant plastid genomes. An analysis of green and red plastid complete genome sequences revealed that 50 core protein-coding genes are common to all algal chloroplasts, with an additional set of 14 genes only associated with red plastids¹⁷. An examination of the mature proteins encoded in the plastid genome of extant green and red phytoplankton species reveals that they do not have significantly different trace elemental requirements. Differences in the elemental composition of the two superfamilies must therefore reflect gene transfers to the host nucleus^{3,4,17}. The retention and expression of these genes within the nuclear genome suggests that they are essential for metabolic processes. If the key selection pressure occurred early in the evolution of the major eukaryotic phytoplankton phyla, as suggested by the phenotypic manifestation of distinct elemental profiles between plastid superfamilies, then we would predict a highly conserved set of primary and secondary gene products that are specific to the plastid type commonly inherited through nuclear genes within a superfamily.

Examining the elemental composition of extant chlorophyll *b*- and *c*-containing dinoflagellates fortuitously provides a test of this 'plastid imprint' hypothesis (Table 1; see also Supplementary Fig. 1). Assuming that the genetic backgrounds of the primordial host cells of dinoflagellates are similar, the hypothesis predicts that the trace elemental composition of dinoflagellates with plastids inherited from the red lineage would cluster more closely with other members of the chlorophyll *c*-containing superfamily, whereas dinoflagellates with green plastids would cluster with the chlorophyll *b*-containing superfamily. On the basis of trace elemental profiles, the three chlorophyll *c*-containing dinoflagellates that we examined, *Prorocentrum minimum*, *Amphidinium carterae* and *Thoracosphaera heimii*, grouped with the red superfamily, whereas the chlorophyll *b*-containing dinoflagellate, *Gymnodinium chlorophorum*, clustered with the green superfamily (Fig. 2). The host cell genetic backgrounds of all four species, however, did not discriminate among the plastid types based on C:N:P stoichiometry (Fig. 1; see also Supplementary Table 1). Our analysis indicates that the bulk elemental composition of phytoplankton is associated with the original heterotrophic host cells' genetic background and that plastid inheritance—but not the contemporary plastid genome—is a key factor determining trace elemental quotas in marine phytoplankton. Differentiation of specific trace elements by the two superfamilies further suggests that the evolutionary history of phytoplankton was strongly influenced by the redox chemistry of the ocean.

Unlike the stoichiometry of major nutrients, that of trace elements in phytoplankton does not resemble the composition of sea water (see Supplementary Fig. 2 insert). This is particularly evident of Fe and Mo, which are, respectively, the most and least abundant essential trace elements in phytoplankton, but the least and most abundant trace elements in the contemporary ocean¹⁸. Indeed, the overall pattern of the relative abundance of trace elements in the ensemble of phytoplankton examined roughly approximates that in the Earth's crust (see Supplementary Fig. 2), where Fe is abundant and Mo very scarce. A similar general trend, observed in higher plants¹⁹, suggests that there is a core elemental requirement of ancestral plastids that has been highly conserved

throughout the evolution of oxygenic photosynthesis. The high requirements for Fe and Mn and low requirement for Mo in all photoautotrophs presumably reflects the high solubility of Fe and Mn and the low solubility of Mo, Cu and other transition elements under the reducing conditions of early Proterozoic oceans²⁰, when oxygenic photosynthesis first appeared²¹. This basic elemental composition helped to underpin the ecological success of the green superfamily throughout the late Proterozoic and the Palaeozoic eras. After the end-Permian extinction, and ocean anoxic events in the Triassic and Early Jurassic periods, the red superfamily rose to ecological prominence²². The relatively low Fe and Mn quotas and high Cd, Co and Mo quotas in the red superfamily suggests an adaptation of the photosynthetic apparatus and ancillary supporting proteins to a highly oxidized surface ocean, presumably in the latter half of the Phanerozoic eon²². Future studies of the trace element composition of sediments throughout the Proterozoic and Phanerozoic eons, coupled with genomic analyses of extant photoautotrophs, will almost certainly help resolve details related to the inheritance of trace element composition of phytoplankton. Our results suggest, however, that historical changes in the redox state of the ocean had a critical role in determining the evolutionary trajectory of this heterogeneous group of photosynthetic organisms. □

Methods

Elemental analyses

All apparatus for medium preparation, culturing, sampling and elemental analyses were prepared using trace-metal clean protocols^{23–25}. Cultures (Table 1) were grown in a modified version of the artificial seawater medium, Aquil²³, for at least six generations before sampling. We lowered the final Fe concentration to 0.36 nM and added Cd (15 nM)²⁴. Cell numbers and sizes were measured with a calibrated Coulter multisizer particle counter to calculate growth rates and cell volumes; specific growth rates were $\geq 90\%$ of the maximum.

Exponentially growing cultures ($n \geq 3$) were collected by filtration onto precombusted Gelman AE glass fibre filters for CHN analyses and on trace-metal clean Poretics polycarbonate filters for analysis by high-resolution inductively coupled plasma mass spectrometry (HR-ICPMS). Samples for CHN analysis were dried in a desiccator at room temperature and assayed with a Carlo Erba elemental analyser. For calcifying species, a duplicate set of filters was fumed for 24 h with concentrated HCl before CHN analysis. Filtrates for HR-ICPMS were washed with either trace-metal clean synthetic ocean water²³ or chelexed 0.42 M NaCl²⁴. Samples and standards for HR-ICPMS were prepared and analysed as described^{24,25}; samples were digested in Teflon vials with 90% HNO₃/10% HF at 120 °C for 4 h. Two isotopes for all elements with multi-isotopes (for example, Fe 56/57, Cu 63/65) were measured with HR-ICPMS to confirm low interference; the results for the more abundant isotope, however, were used in our analyses because of the superior counting statistics. Elemental concentrations were corrected for blanks and normalized on a per cell basis for statistical analyses.

Statistical analyses

To determine whether the average elemental quotas were different between phyla and superfamilies, we performed a series of *t*-tests and ANOVAs with a nested design. The *P*-values for rejecting the null hypothesis (equal means in two different groups) were adjusted for multiple comparisons²⁶. Patterns in trace elemental composition were further examined using principal component analysis^{12,27}. The analysis was performed on Fe, Mn, Zn, Cu, Co and Cd concentrations, normalized to phosphorus and log transformed to homogenize the variances. Each species was represented as a point in six-dimensional space (one dimension for each element). The analysis was performed on the basis of correlations rather than co-variances, giving equal weight to each ratio as opposed to giving more weight to the elements with the larger ratios. The results were projected onto a plane defined by the first two principal components. Projections of a unit vector pointing along each ratio axis were calculated, making it possible to interpret how a change in position on the plot corresponded to a change in the ratio.

Received 14 April; accepted 21 July 2003; doi:10.1038/nature01953.

1. Haeckel, E. *Plankton-Studien. Vergleichende Untersuchungen über die Bedeutung und Zusammensetzung der pelagischen Fauna und Flora* (G. Fischer, Jena, 1890).
2. Delwiche, C. F. Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* **154**, S164–S177 (1999).
3. Martin, W. *et al.* Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99**, 12246–12251 (2002).
4. Palmer, J. D. The symbiotic birth and spread of plastids: how many times and howdunit? *J. Phycol.* **39**, 4–11 (2003).
5. Falkowski, P. G. & Raven, J. A. *Aquatic Photosynthesis* (Blackwell, Malden, MA, 1997).
6. Redfield, A. C. in *James Johnstone Memorial Volume* (ed. Daniel, R. J.) 176–192 (Liverpool Univ. Press, Liverpool, 1934).

7. Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature* **387**, 272–275 (1997).
8. Sverdrup, H. U., Johnson, M. W. & Fleming, R. H. *The Oceans* (Prentice-Hall, New Jersey, 1942).
9. Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: The Biology of the Elements from Molecules to the Biosphere* (Princeton Univ. Press, Princeton, 2002).
10. Falkowski, P. G. Rationalizing elemental ratios in unicellular algae. *J. Phycol.* **36**, 3–6 (2000).
11. Geider, R. J. & LaRoche, J. Redfield revisited: variability of C:N:P in marine microalgae and its biochemical basis. *Eur. J. Phycol.* **37**, 1–17 (2002).
12. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S-PLUS* (Springer, New York, 1999).
13. Raven, J. A., Evans, M. C. W. & Korb, R. E. The role of trace metals in photosynthetic electron transport in O₂-evolving organisms. *Photosynth. Res.* **60**, 111–149 (1999).
14. Falkowski, P. G., Owens, T. G., Ley, A. C. & Mauzerall, D. C. Effects of growth irradiance levels on the ratio of reaction centers in two species of marine phytoplankton. *Plant Physiol.* **68**, 969–973 (1981).
15. Price, N. M. & Morel, F. M. M. Cadmium and cobalt substitution for zinc in a marine diatom. *Nature* **344**, 658–660 (1990).
16. Lee, J. G. & Morel, F. M. M. Replacement of zinc by cadmium in marine phytoplankton. *Mar. Ecol. Prog. Ser.* **127**, 305–309 (1995).
17. Grzebyk, D., Schofield, O., Vetrani, C. & Falkowski, P. G. The Mesozoic radiation of eukaryotic algae: The portable plastid hypothesis. *J. Phycol.* **39**, 259–267 (2003).
18. Whitfield, M. Interactions between phytoplankton and trace metals in the ocean. *Adv. Mar. Biol.* **41**, 3–128 (2001).
19. Williams, R. J. P. & da Silva, J. J. R. F. *The Natural Selection of the Chemical Elements* (Clarendon, Oxford, 1996).
20. Andar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and the evolution: A bioinorganic bridge? *Science* **297**, 1137–1142 (2002).
21. Des Marais, D. J. When did photosynthesis emerge on Earth? *Science* **289**, 1703–1705 (2000).
22. Falkowski, P. G. et al. in *Coccolithophores: Molecular Processes to Global Impact* (eds Thierstein, H. & Young, J.) (Springer, Berlin, in the press).
23. Price, N. M. et al. Preparation and chemistry of the artificial algal culture medium Aquil. *Biol. Oceanogr.* **6**, 443–461 (1988/89).
24. Ho, T.-Y. et al. The elemental composition of some marine phytoplankton. *J. Phycol.* (submitted).
25. Cullen, J. T., Field, T. S. & Sherrell, R. M. The determination of trace elements in filtered suspended marine particulate material by sector field HR-ICP-MS. *J. Anal. At. Spectrom.* **16**, 1307–1312 (2001).
26. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
27. Ihaka, R. & Gentleman, R. R. Language for data analysis and graphics. *J. Comp. Graph. Stat.* **5**, 299–314 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Wyman, C. Fuller, P. Field and R. Sherrell for assisting us with the elemental analysis, and L. Hedin, R. Sherrell and J. Raven for comments. This work was supported by the National Science Foundation 'Evolution and Radiation of Eukaryotic Phytoplankton Taxa' (EREUPT) Biocomplexity Program (Rutgers University) and the Centre for Environmental Bioinorganic Chemistry at the Princeton Environmental Institute (Princeton University).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.Q. (aquigg@imcs.rutgers.edu) or P.G.F. (falko@imcs.rutgers.edu).

Predicted recurrences of mass coral mortality in the Indian Ocean

Charles R. C. Sheppard

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

In 1998, more than 90% of shallow corals were killed on most Indian Ocean reefs¹. High sea surface temperature (SST) was a primary cause^{2,3}, acting directly or by interacting with other factors^{3–7}. Mean SSTs have been forecast to rise above the 1998 values in a few decades^{2,3}; however, forecast SSTs rarely flow seamlessly from historical data, or may show erroneous seasonal oscillations, precluding an accurate prediction of when lethal SSTs will recur. Differential acclimation by corals in different places complicates this further^{3,7,8}. Here I scale forecast SSTs at 33

Indian Ocean sites where most shallow corals died in 1998 (ref. 1) to identify geographical patterns in the timing of probable repeat occurrences. Reefs located 10–15° south will be affected every 5 years by 2010–2025. North and south from this, dates recede in a pattern not directly related to present SSTs; paradoxically, some of the warmest sites may be affected last. Temperatures lethal to corals vary in this region by 6 °C, and acclimation of a modest 2 °C by corals could prolong their survival by nearly 100 years.

Timing of recovery from the 1998 massive coral mortality on Indian Ocean reefs (refs 1, 2, 6 and Fig. 1) and how frequently rising SSTs will cause repeat mortalities are issues of practical urgency for many countries because of the high value of reefs to shoreline protection, biodiversity, protein supply and tourism^{6,9,10}. Raw, modelled SSTs cross supposed thresholds of coral bleaching in a few decades², but scaling problems in forecast data, coral acclimation^{3,7,8} and different absolute SSTs and rates of SST rise vary markedly between sites, which greatly affects estimates of when rising temperatures will reach values that proved lethal before. Exact dates remain unattainable, but a probability approach proves very revealing in terms of both timing and geographical pattern. Using the Indian Ocean, whose reefs were worst affected in the warm SSTs of 1998, I have 'blended' forecast SSTs seamlessly onto historical SST data, with appropriately scaled forecast seasonal cycles, for 33 sites.

Historical SST data from 1871–1999 from the HadISST1 data set^{11,12} were combined with surface ('skin') temperature from 1950–2099 from the HadCM3 model for each site; the latter equates with SST (see <http://www.cru.uea.ac.uk/cru/info/modelcc/>). Both series are monthly; HadISST1 cells are 1° latitude and longitude, whereas the HadCM3 cells used are 2.5 × 3.75° (<http://www.cru.uea.ac.uk/cru/info/modelcc/>). HadISST1 comprises the latest historical data¹¹ and HadCM3 is the most recent 'coupled' model from the Hadley

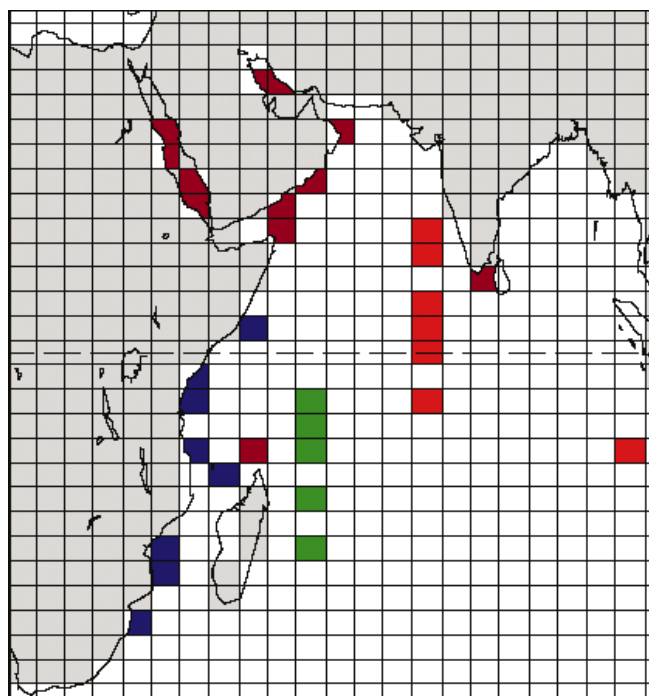


Figure 1 Coral reef sites in the western Indian Ocean where 1998 SSTs caused mass coral mortality. The grid comprises HadCM3 cells. Blue, green and red filled cells show the transects studied; colours match the curves in Fig. 4. Brown shows additional sites examined (Arabian region, Aldabra and Sri Lanka) that do not form coherent transects (Sri Lanka is located near the central chain of atolls but is a 'high island' with reefs of very different character; Cocos Keeling, the eastern-most site, is a typical atoll and is included in the central atolls group.)

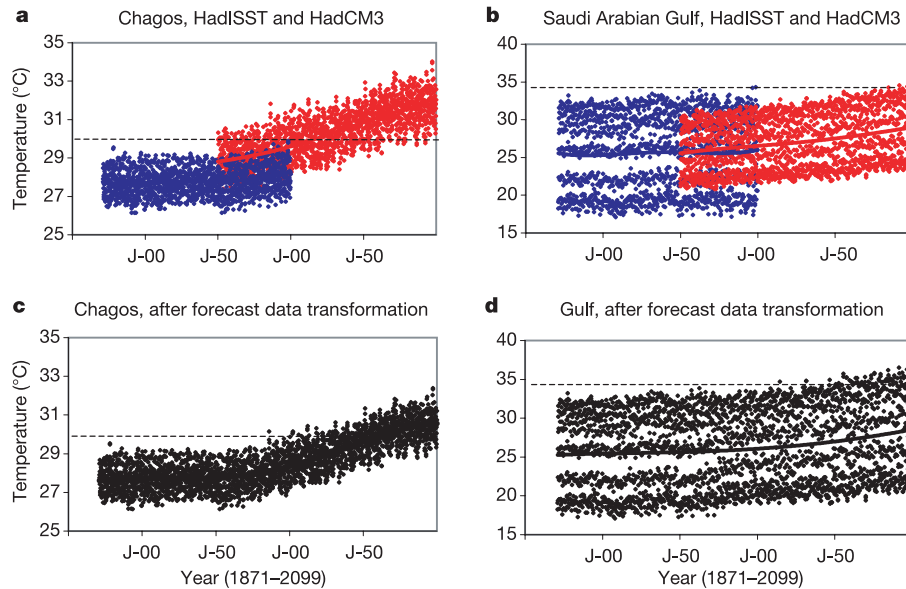


Figure 2 Historical and forecast SST data from two sites before and after the transformations. Historical data (HadISST1) are blue, and forecast data (HadCM3, IS92a) are red. Black points are data after transformation. **a, c**, Typical oceanic site of Chagos

archipelago. **b, d**, Extreme continental site of Saudi Arabian Gulf. Dotted horizontal lines indicate SST values that killed corals in 1998. Curves of best fit are Excel's fourth-order polynomials.

Centre¹³. Other data sets can be substituted with minimal changes to the pattern described (Supplementary Information). With regard to warming schemes, more than 40 exist today; the one chosen is the thoroughly tested IS92a scheme¹⁴, which follows a median path^{13,14} and has been most used for inter-comparisons (Supplementary Information).

'Raw' SST data require transformation before use (Fig. 2). In typical oceanic atolls (Chagos), historical and forecast series are discontinuous by more than 1.5 °C. In an exceptionally variable site (Saudi Arabian Gulf), measured SST amplitude is the highest known for coral reefs (15 °C) but its forecast SST series shows an erroneously small seasonality, which would mislead because it is the annual high SST that causes mortality¹⁵.

From each SST series (Methods), the probability of repeat critical SSTs can be determined. Four sites illustrate this (Fig. 3): those noted above plus those with the fastest and slowest rates of rise among these 33 sites. These curves integrate: the absolute

SST at a site, its rate of rise, and the temperature that was lethal to more than 90% of the shallow corals there in 1998, which is a function of acclimation. The coral species are all from the same western Indian Ocean subset of fauna, with relatively few endemic exceptions^{16,17}.

Mortality in 1998 was triggered by SST rises lasting as short as the warmest month, although warming lasted 3 months in many areas^{1,2,18–22}. As it is not yet known exactly how much warming triggers bleaching leading to mortality (or for how long), probability of recurrence curves are computed with warmest month, warmest 3 months and averaged warmest quarter. To compare these across sites, a uniform 'extinction date' is required. Although almost any point will suffice to show the geographical pattern, the date chosen is the year when there is a 0.2 probability of the warmest month (or warmest 3 months or averaged warmest quarter) hitting the 1998 value. The value 0.2 is ecologically fairly realistic: most corals do not mature until 5 years old^{2,23} and today, 5 years after the

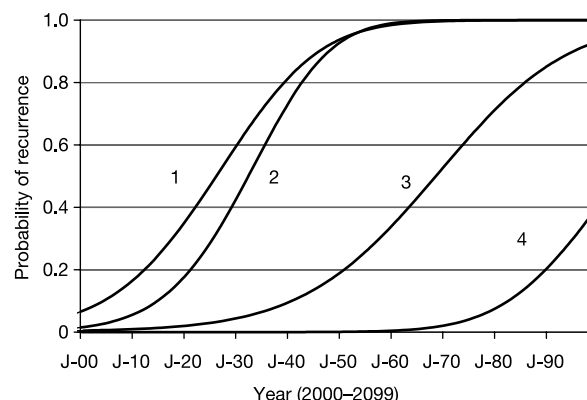


Figure 3 Probabilities of the warmest months of four sites reaching the lethal 1998 temperatures over time. Middle curves are sites used in Fig. 2; outer curves are two extremes. 1, Comores (African coast transect); 2, Chagos; 3, Saudi Gulf; 4, Minicoy, North

Lakshadweep. The warmest month was March in Comoros and Chagos, May in North Lakshadweep, and September in Saudi Arabian Gulf.

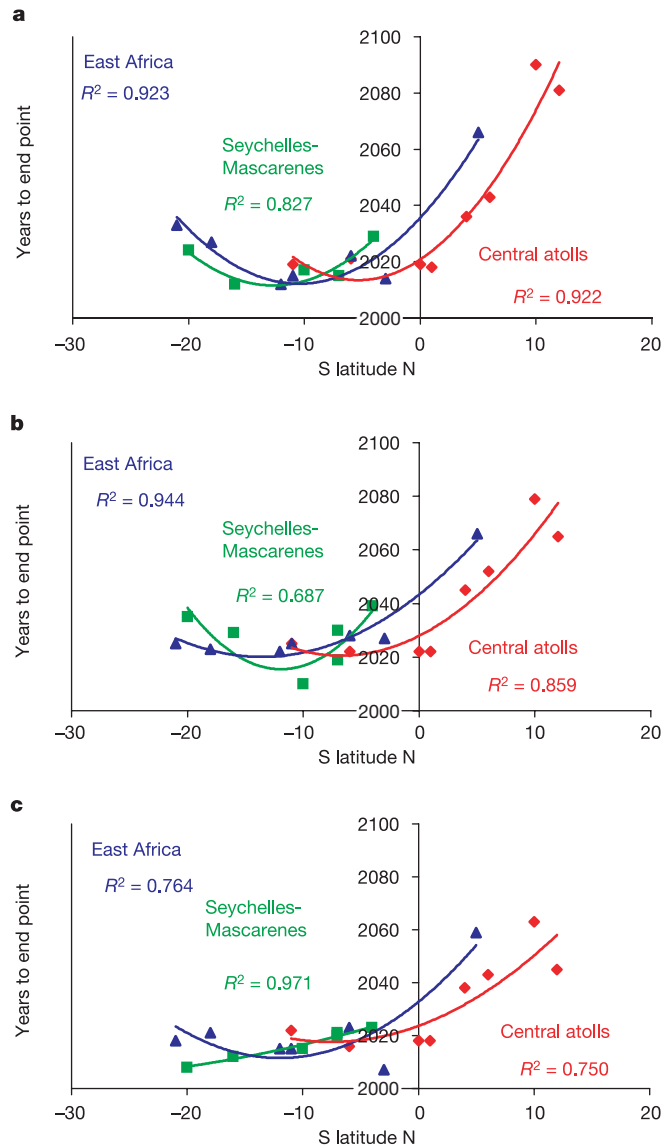


Figure 4 'Extinction dates' plotted for coral reef sites on the three transects, showing the date when a probability of 0.2 is reached. All curves are significant fits. **a**, Warmest month. **b**, Warmest 3 months. **c**, Warmest quarter. Colours match sites of Fig. 1. If the eastern Cocos Keeling atoll is removed from the central atoll transect, R^2 values for that transect remain almost unchanged (0.921, 0.868 and 0.780, respectively). If the continental reefs of Sri Lanka are added to that transect, R^2 values drop slightly (0.742, 0.833 and 0.747, respectively) but are still significant ($P < 0.01$).

1998 event, most of these sites have recovered only marginally¹⁹, with coral cover rising in shallow water from 1–2% immediately after 1998 to about 3–5% today, as compared with pre-1998 values of 40–75% (refs 19–22).

Whether warmest month, warmest 3-month period or averaged warmest quarter values are used, earliest vulnerability will occur between 10–15° south (Fig. 4) between 2010 and 2025 along all three north–south transects. These 'extinction dates' recede southwards, but also recede northwards, initially towards the equator. The Arabian sites have a confused pattern owing partly to a "pseudo-high latitude effect"²⁴ caused by cold summer upwelling in the Arabian Sea, such that some sites with highest temperatures have the most distant extinction dates; in others, upwelling precluded determination of the 'extinction date' (Supplementary Information). Many Arabian corals annually survive temperatures that

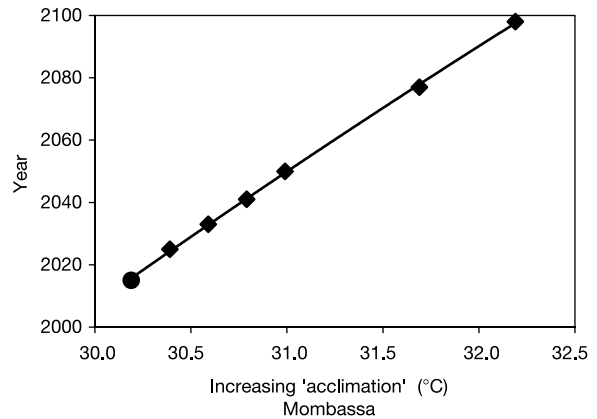


Figure 5 Recession of time to extinction date with imagined acclimation of corals by up to 2°C with Kenya as the example. Dot is the existing situation, diamonds are the extinction date given coral 'acclimation'.

killed the same species elsewhere in 1998 (refs 15, 19, 20). The northern Red Sea remains relatively unaffected but, even with their marked temperature acclimation, most corals in the Arabian Gulf were killed by the 1998 peak SSTs. The fact that most sites between 0° and 15° south will have a 1 in 5 probability annually of suffering a month as warm as that of 1998 within 10–15 years means that several of the world's poorest countries, for which reefs provide essential resources^{6,9,10}, will be affected soonest.

A modest acclimation or adaptation by corals^{7,8} would greatly prolong time before their 'extinction date'. In 1998, lethal SSTs varied by 5°C (from <29°C to >34°C) depending on location. By raising the SST presumed to be lethal at a site by 2°C, nearly a century will be gained (Fig. 5). A value of 2°C seems modest compared with that already achieved by today's Arabian corals^{15,17}, but the latter had millennia to acclimate. In these 33 sites today, total coral cover has improved little since 1998 (refs 18–20). The few decades before probable recurrence of lethal SST values may not be sufficient for recovery to become well established.

Diverse physiological and pathological factors are triggered by a rise in SST. Most act with, or are triggered by, temperature⁷. Some physical factors such as increased cloudiness⁴ and water exchange⁵ oppose these effects. SST should not be taken as the only important factor, although it is at worst a quantitative, measurable surrogate. □

Methods

Scaling of forecast data series

SSTs forecast from climate models rarely flow seamlessly from historical series, and errors in forecast seasonal amplitudes further prevent accurate estimation of when lethal mortalities might recur. The first transformation simply adjusts each forecast data series by the mean difference in values in the overlapping data between 1950 and 1999 ($n = 600$ months).

The second transformation scaled the seasonal amplitude of each forecast series to match that of each site's historical data. Fourth-order polynomial fits were computed for each historical series. Fourth order was chosen by extensive trial and error; orders up to fourth significantly increased R^2 , whereas higher orders added no significant further improvement. Predicted values were subtracted from their corresponding monthly data points to obtain series of residuals. This was repeated for forecast data at each site. Correlations between residuals of the historical and forecast SSTs at each site in overlapping years (1950–1999) are always highly significant; in the examples plotted, (Fig. 2), $r = 0.973$ for the Gulf and $r = 0.762$ for Chagos ($n = 600$).

By using the 'normdist' function of Excel (Microsoft), residuals of forecast series were expressed as standard deviations. By substituting the standard deviation of the historical data residuals in place of that of the forecast data, Excel's 'norminv' function was then used to compute forecast temperature residuals whose annual oscillation matches in magnitude that of the historical series. Adding these scaled residuals back to the polynomial curves gives, for every site, an SST series (1871–2099) with no disjunction and, where they overlap, the same seasonal amplitude.

Standard deviations of HadISST1 residuals are stable with time¹¹. HadCM3 residuals increase in amplitude by 3–25% with time, reflecting the climate model's increasing

uncertainty into the future, and this is carried through in the transformed series. Other historical SST series could be substituted; the changes to the pattern are minor to negligible (Supplementary Information).

Computing probabilities of recurrence

Subsets of data were extracted for warmest months, three warmest months and averaged warmest quarters of each year. Residuals of all but one (Alphonse atoll) warmest month series have normal distributions (Kolmogorov–Smirnov tests). Warmest quarters' residuals are also normally distributed in all sites except one (granitic Seychelles). As time proceeds, the difference between the lethal 1998 SST value (also expressed as a residual) and the normally distributed population of SSTs decreases. For each month, '1 – normdist' determines the probability that each site's lethal temperature is part of the site's population of temperatures. This yields probability curves of repeat recurrences of the peak temperature of 1998.

In the warmest 3-month data sets, residuals in only half of the sites have normal distributions (they lack extended 'tails'). For these, 'bootstrap tests' were used instead of the normdist function to compute probability; probability was the number of residuals in the whole data set with a value greater than the test value, divided by the total number. Curves almost exactly match those obtained by the normdist method. This test was also used for the north Seychelles site for the quarterly series to extend that transect northwards; the test differed by less than 1 year from that obtained with the normdist function at that site.

'Lethal' SSTs and timing

Peak SSTs ranged from February in the south to September in the northwest. For 27 sites, the warmest quarter was the peak month with the preceding and the following months. For the other six sites, it was the peak month with the two preceding months. For the warmest month and 3-month tests, the test SST value was the warmest 1998 HadISST1 temperature. For the warmest quarter test, the average SST of the warmest 3 months was used. These temperatures were generally only less than 0.2 °C warmer than any earlier recorded temperatures at that site.

Received 13 May; accepted 5 August 2003; doi:10.1038/nature01987.

1. Wilkinson, C. R. in *Seas at the Millennium, an Environmental Evaluation* vol. 3 (ed. Sheppard, C.) 43–57 (Elsevier, Amsterdam, 2000).
2. Hoegh-Guldberg, O. Climate change, coral bleaching and the future of the world's coral reefs. *Mar. Freshwater Res.* **50**, 839–866 (1999).
3. Pittcock, A. B. Coral reefs and environmental change: Adaptation to what? *Am. Zool.* **39**, 10–29 (1999).
4. Turner, J., et al. in *Coral Reef Degradation in the Indian Ocean, Status Report 2000* (eds Souter, D., Obura, D. & Linden, O.) 94–107 (CORDIO, Stockholm, 2000).
5. Nakamura, T. & Woelke, R. Water-flow rates and passive diffusion partially explain differential survival during the 1998 bleaching event. *Mar. Ecol. Prog. Ser.* **212**, 301–304 (2001).
6. Sheppard, C. R. C. The main issues affecting coasts of the Indian and western Pacific Oceans: a meta-analysis from seas at the millennium. *Mar. Pollut. Bull.* **42**, 1199–1207 (2001).
7. Douglas, A. W. Coral bleaching—how and why? *Mar. Pollut. Bull.* **46**, 385–392 (2003).
8. Gates, R. D. & Edmunds, P. J. The physiological mechanisms of acclimatisation in tropical reef corals. *Am. Zool.* **39**, 30–43 (1999).
9. Wilkinson, C. R. et al. Ecological and socio-economic impacts of 1998 coral mortality in the Indian Ocean: an ENSO (El Niño–Southern Oscillation) impact and a warning of future change? *Ambio* **28**, 188–196 (1999).
10. Cesar, H. S. J. (ed.) *Economics of Coral Reefs* 244 (CORDIO, Kalmar, Sweden, 2000).
11. Rayner, N. A. et al. Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmos.* (in the press).
12. Sheppard, C. R. C. & Rayner, N. Utility of the Hadley Centre sea ice and surface temperature data set (HadISST1) in two widely contrasting coral reef areas. *Mar. Pollut. Bull.* **44**, 303–308 (2002).
13. McAvaney, B. J. et al. *Climate Change 2001: The Scientific Basis. Intergovernmental Panel on Climate Change* (ed. Houghton, J. T. et al.) 471–523 (Cambridge Univ. Press, Cambridge, UK, 2001).
14. Cubasch, U. et al. *Climate Change 2001: The Scientific Basis. Intergovernmental Panel on Climate Change* (ed. Houghton, J. T. et al.) 525–582 (Cambridge University Press, Cambridge, UK, 2001).
15. Sheppard, C. R. C. & Loughland, R. Coral mortality and recovery in response to increasing temperature in the southern Arabian Gulf. *Aquat. Ecosyst. Health Mgmt* **5**, 395–402 (2002).
16. Veron, J. E. N. *Coral Reefs of the World* vols 1–3 (Australian Institute of Marine Science, Townsville, Australia, 2000).
17. Sheppard, C. R. C. & Sheppard, A. L. S. Corals and coral communities of Arabia. *Fauna Saudi Arabia* **12**, 7–192 (1991).
18. Sheppard, C. R. C., Spalding, M., Bradshaw, C. & Wilson, S. Erosion vs. recovery of coral reefs after 1998 El Niño: Chagos reefs, Indian Ocean. *Ambio* **31**, 40–48 (2002).
19. Linden, O., Souter, D., Willemsson, D. & Obura, D. *Coral Reef Degradation in the Indian Ocean, Status Report 2002* 284 (CORDIO, Kalmar, Sweden, 2002).
20. Riegl, B. Effects of the 1996 and 1998 positive sea-surface temperature anomalies on corals, coral diseases and fish in the Arabian Gulf (Dubai, UAE). *Mar. Biol.* **140**, 29–40 (2002).
21. McClanahan, T. R. Bleaching damage and recovery potential of Maldivian coral reefs. *Mar. Pollut. Bull.* **40**, 587–597 (2000).
22. Spencer, T., Teleki, K. A., Bradshaw, C. & Spalding, M. D. Coral bleaching in the southern Seychelles during the 1997–1998 Indian Ocean warm event. *Mar. Pollut. Bull.* **40**, 569–586 (2000).
23. Harrison, P. L. & Wallace, C. C. in *Coral Reefs* (ed. Dubinsky, Z.) 133–206 (Elsevier, Amsterdam, 1990).
24. Sheppard, C. R. C. & Salm, R. V. Reefs and corals of Oman, with a description of a new species of coral (*Scleractinia, Acanthastrea*). *J. Nat. Hist.* **22**, 263–279 (1988).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The HadCM3 data were provided by the Hadley Centre for Climate Research through D. Viner, who also provided information on the data's characteristics. I thank M. Keeling and G. Medley for advice on analyses; N. Rayner of the Hadley Centre for information on the HadISST1 data and for communicating results before publication; and O. Langmead and A. Edwards for assistance with data extraction.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.C.S. (csheppard@bio.warwick.ac.uk).

Monkeys reject unequal pay

Sarah F. Brosnan & Frans B. M. de Waal

Living Links, Yerkes National Primate Research Center, Emory University, Atlanta, Georgia 30329, USA

During the evolution of cooperation it may have become critical for individuals to compare their own efforts and pay-offs with those of others. Negative reactions may occur when expectations are violated. One theory proposes that aversion to inequity can explain human cooperation within the bounds of the rational choice model¹, and may in fact be more inclusive than previous explanations^{2–8}. Although there exists substantial cultural variation in its particulars, this 'sense of fairness' is probably a human universal^{9,10} that has been shown to prevail in a wide variety of circumstances^{11–13}. However, we are not the only cooperative animals¹⁴, hence inequity aversion may not be uniquely human. Many highly cooperative nonhuman species seem guided by a set of expectations about the outcome of cooperation and the division of resources^{15,16}. Here we demonstrate that a nonhuman primate, the brown capuchin monkey (*Cebus apella*), responds negatively to unequal reward distribution in exchanges with a human experimenter. Monkeys refused to participate if they witnessed a conspecific obtain a more attractive reward for equal effort, an effect amplified if the partner received such a reward without any effort at all. These reactions support an early evolutionary origin of inequity aversion.

In preliminary studies, two conditions were used: 'equality', in which two monkeys exchanged tokens with a human experimenter

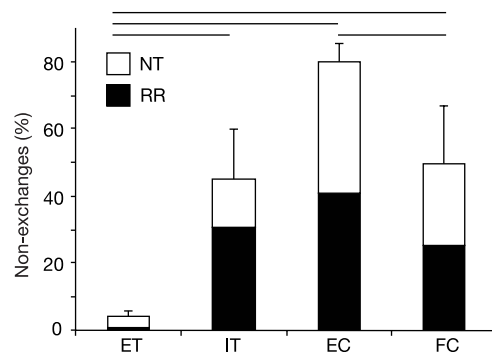


Figure 1 Mean percentage $\pm$ s.e.m. of failures to exchange for females across the four test types. Black bars (RR) represent the proportion of non-exchanges due to refusals to accept the reward; white bars (NT) represent those due to refusals to return the token. s.e.m. is for combined non-exchanges. Lines indicate significant differences between conditions (Tukey's multiple comparisons). ET, equality test; IT, inequality test; EC, effort control; FC, food control.

to receive cucumber, and 'inequality', in which one monkey exchanged for cucumber and its partner for grape, a more favoured food. Whereas in previous tests males and females had been equally reliable exchangers, only females reacted differently to the two conditions. Compared with equality tests, females receiving the less favoured reward in inequality tests were less willing to exchange, whereas males showed no such effect. Our limited sample size did not allow a conclusive comparison of the sexes, but independent evidence indicates that capuchin females pay closer attention than males to the value of exchanged goods and services^{17,18}.

The study reported here concerns only five females, which were again subjected to the equality and inequality conditions plus two new controls: 'effort controls', in which a grape was simply handed to the partner by the experimenter (no exchange) followed by the subject herself exchanging for cucumber, and 'food controls' in which, in the absence of a partner, the subject witnessed a grape being placed in the location where the partner normally sat, after which the subject herself exchanged for cucumber. For this test, grapes accumulated every trial, as it would have been too disruptive to remove them. We measured the monkeys' rate of and latency to successful exchange. Although available rewards were visible to both monkeys, neither was shown which reward they would receive before successfully returning the token. We divided incomplete exchanges into two categories: (1) failure to hand back the token (no token, NT), and (2) failure to accept or eat the proffered reward (reject reward, RR). Both kinds of incomplete exchanges often involved active rejection, such as tossing the token or reward out of the test chamber.

Despite the small number of subjects, the overall exchange tendency varied significantly across the four conditions in a manner indicating that the presence of high-value rewards reduced the tendency to exchange for low-value rewards ($F_{3,16} = 25.78$, $P < 0.001$; Fig. 1). This effect was enhanced by a difference in effort between partners (Tukey's multiple comparisons comparing effort controls with food controls, $P < 0.05$; and effort controls with equality tests, $P < 0.05$). Thus, the strongest increase in refusal to exchange occurred if a partner received better rewards without any effort.

Exchange behaviour may change over the course of a test. Each subject received two tests of each condition, each with 25 trials (exchanges). Failed exchanges (NT plus RR, see above) might increase over consecutive trials if subjects did not immediately recognize that they were receiving a lesser reward, but learned over time. Conversely, failed exchanges might decrease if subjects gradually 'settled' for the lesser reward, seeing that no higher-value reward was forthcoming. We tested these contrasting predictions by comparing each subject's responses during the first 15 versus the last 10 trials per tests. We ran the required exact Wilcoxon signed ranks

test, which given our sample size could not reach a P -value below 0.06. For equality, inequality and effort control conditions, non-exchanges increased over the course of tests, reaching the minimum P -value only for the effort controls (equality: $T = 9$, $n = 4$, $P = 0.25$; inequality: $T = 13$, $n = 5$, $P = 0.18$; effort control $T = 15$, $n = 5$, $P = 0.06$; Fig. 2). For the food controls, in contrast, non-exchanges decreased over the course of tests ($T = -15$, $n = 5$, $P = 0.06$). This last condition differed from the others in that no other individual was present. Perhaps this made adjustments to low-value rewards easier.

Failure to exchange was roughly equally divided between no token exchange (NT, 45.4%) and no food acceptance (RR, 54.6%). Each measure separately showed significant variation across the four conditions (NT, $F_{3,16} = 8.43$, $P = 0.001$; RR, $F_{3,16} = 7.73$, $P = 0.002$; Fig. 1), mainly due to a significant difference in exchange between the equality test and the other three conditions. The general tendencies across the four conditions were the same for NT and RR (Fig. 1), indicating that conditions affected these responses similarly.

Failure to hand back a received token (NT) is a highly unusual response in our trained capuchins: in two years of bartering, such failures occurred in less than 5% of trials, as also seen in the equality test. The marked increase in failure to exchange in individuals receiving the lower-value reward in the inequality test and the two control conditions cannot be explained by the absence of positive reinforcement, as rewards continued to be cucumber, an accepted reward in the equality test. Even more curious than a drop in the conditioned response rate was the second manner in which exchanges failed: refusal to accept or consume the reward (RR). In doing so, subjects forfeited a directly accessible food that they readily accept and consume under almost any other set of circumstances. One possible explanation is that reward rejections relate to violated expectations, in which monkeys forego a low-value reward if a high-value one is anticipated¹⁹. On the basis of her own reward history, however, there would seem no reason for a subject receiving cucumber to expect anything else during the same test. If expect-

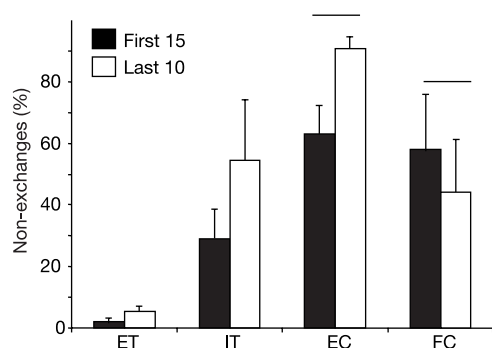


Figure 2 Mean percentage $\pm$ s.e.m. of failures to exchange in the first 15 trials (black bars) versus the last 10 trials per test (white bars). Lines indicate differences at $P = 0.06$ (exact Wilcoxon signed ranks test). ET, equality test; IT, inequality test; EC, effort control; FC, food control.

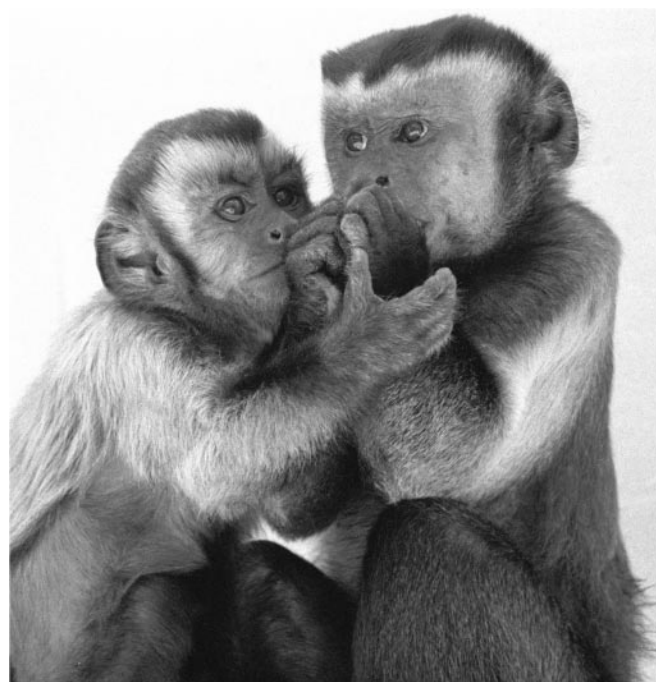


Figure 3 A juvenile capuchin monkey shows cheek-to-cheek begging to an eating adult male, cupping his hand next to the adult's food in solicitation. This primate is exceptionally tolerant and readily shares food, which may be a precondition for the reported reactions to reward division.

tations do have a role, therefore, they must be based on the visible presence of inaccessible high-value rewards. The general increase in refusals to exchange in the course of testing supports this argument. Food controls, in which a partner was absent, differed by showing a decrease in refusals. This suggests that expectations are based on seeing a partner receive and/or eat high-value rewards rather than the mere presence of such rewards.

Finally, when measuring the latency to exchange for all successfully completed exchanges (that is not including NT non-exchanges), we found significant variation across the four conditions ($F_{3,16} = 321.33$, $P < 0.001$), and significant variation between all conditions (Tukey's multiple comparisons). The capuchins completed exchanges most quickly in inequality tests, and more slowly in the control tests than in equality tests. This result is hard to interpret, but precludes the possibility that the mere presence of grapes either excited subjects into more rapid exchange or diverted their attention, thus slowing exchange. Grapes were visible under all conditions except equality tests.

People judge fairness based both on the distribution of gains and on the possible alternatives to a given outcome^{20,21}. Capuchin monkeys, too, seem to measure reward in relative terms, comparing their own rewards with those available, and their own efforts with those of others. They respond negatively to previously acceptable rewards if a partner gets a better deal. Although our data cannot elucidate the precise motivations underlying these responses, one possibility is that monkeys, similarly to humans, are guided by social emotions. These emotions, known as 'passions' by economists, guide human reactions to the efforts, gains, losses and attitudes of others^{22–25}. Clearly if these reactions evolved to promote long-term human cooperation²⁶, they may exist in other animals as well²⁷. It has been proposed that nonhuman primates are guided by species-typical expectations "about the way in which oneself (or others) should be treated and how resources should be divided"¹⁵. As opposed to primates marked by despotic hierarchies, tolerant species with well-developed food sharing and cooperation, such as capuchins (Fig. 3)^{16,18,28–30}, may hold emotionally charged expectations about reward distribution and social exchange that lead them to dislike inequality. □

Methods

Five adult female and five male (three adult and two subadult) capuchin monkeys from two long-term, stable social groups were initially tested, although only the females completed all testing. Experiments were carried out in a familiar test chamber, which was divided by a mesh partition into two equally sized (36 × 60 × 60 cm) compartments. For testing, the subject and partner were enclosed in these adjoining compartments and had visual, vocal and limited tactile contact (that is, both monkeys of an experimental pair were close enough to clearly observe the exchanges and rewards of the other). To minimize distractions from their group, an opaque panel backed the test chamber, allowing some vocal but no visual or tactile contact. For details about the testing facility, see ref. 18.

For exchange, the subject was given a token that could immediately be returned to the experimenter for a food reward. The experimenter (S.F.B.) stood before the monkey with the left hand outstretched in a palm-up begging gesture, approximately 5 cm above the floor of the test chamber and 2 cm from the mesh, and with the right hand in the laboratory coat pocket. No other cues were given to encourage the monkey, and no reward was shown before correct exchange. The monkey had 60 s to place the token into the palm of the experimenter's outstretched hand. Throwing the token at the experimenter or out of the test chamber did not count as an exchange. After a successful return, the experimenter lifted the correct reward from a transparent bowl visible to both monkeys and gave it to the exchanger. In cases of failure to exchange, no reward was shown. Exchange interactions were typically completed in about 5 s, and all subjects exchanged successfully in 95% or more of the baseline tests. All subjects had at least 2 yr experience with the exchange scheme, but none had been used previously in any similar study.

Tokens were small granite rocks (2–3 cm in diameter). The lower-value food item was one-quarter of a cucumber slice and the higher-value food item was a grape. In dichotomous choice tests, grapes were preferred at least 90% of the time to cucumber by all capuchins, but all capuchins readily exchanged for cucumber in the absence of other rewards (for example, equality test). Food preferences did not change throughout the testing period.

Capuchins were paired with a same-sex partner who remained the same throughout the study. The exception was the food control test, in which subjects were tested alone. Monkeys were never used in more than one test per day. Each condition was tested twice, on different days, with each test consisting of 50 alternating trials (25 per

individual) beginning with the partner. Each monkey had 60 s to exchange, and after either 60 s or the completion of an exchange, whichever came first, the other monkey's exchange commenced. No break, other than that necessary for the exchanger to prepare (5–10 s), was added between trials, to assure that subjects remembered the previous exchange. The presence of a partner consuming a reward did not visibly affect the subjects' behaviour during exchange. All tests consisted of trials of one condition only. Each capuchin was tested on all four conditions: equality, inequality, effort control and food control.

Failures to exchange were divided into refusals to return the token and refusals to accept the reward. Data were collected during testing by a second experimenter, who remained the same throughout the study, with supplementary data obtained from videotapes. All experiments were videotaped using a Canon GL-1 digital camcorder, time stamped to 100th of a second. Friedman's tests were used to compare individuals' mean failure to exchange across test types, and pair-wise comparisons were conducted using Tukey's multiple comparisons. When comparing exchange rate in the first and second half of tests, exact Wilcoxon signed ranks tests were used. All reported *P*-values are two-tailed.

Received 14 May; accepted 23 July 2003; doi:10.1038/nature01963.

1. Fehr, E. & Schmidt, K. M. A theory of fairness, competition, and cooperation. *Quart. J. Econ.* **114**, 817–868 (1999).
2. Andreoni, J. Cooperation in public-goods experiments: kindness or confusion? *Am. Econ. Rev.* **85**, 891–904 (1995).
3. Fehr, E. & Gächter, S. The puzzle of human cooperation: reply. *Nature* **421**, 912 (2003).
4. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
5. Johnson, D. D. P., Stopka, P. & Knights, S. The puzzle of human cooperation. *Nature* **421**, 911–912 (2003).
6. Milinski, M., Semmann, D. & Krambeck, H.-J. Reputation helps solve the 'tragedy of the commons'. *Nature* **415**, 424–426 (2002).
7. Roberts, G. & Sherratt, T. N. Development of cooperative relationships through increasing investment. *Nature* **394**, 175–179 (1998).
8. Wenekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **200**, 850–852 (2000).
9. Henrich, J. et al. In search of Homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* **91**, 73–78 (2001).
10. Henrich, J. Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *Am. Econ. Rev.* **90**, 973–979 (2000).
11. Fehr, E. & Rockenbach, B. Detrimental effects of sanctions on human altruism. *Nature* **422**, 137–140 (2003).
12. Kahneman, D., Knetsch, J. L. & Thaler, R. Fairness as a constraint on profit seeking: entitlements in the market. *Am. Econ. Rev.* **76**, 728–741 (1986).
13. Zizzo, D. J. & Oswald, A. Are people willing to pay to reduce other's incomes? *Annales d'Econ. Stat.* **63–64**, 39–62 (2001).
14. Dugatkin, L. A. *Cooperation Among Animals: An Evolutionary Perspective* (Oxford Univ. Press, New York, 1997).
15. de Waal, F. B. M. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals* (Harvard Univ. Press, Cambridge, Massachusetts, 1996).
16. de Waal, F. B. M. & Davis, J. M. Capuchin cognitive ecology: cooperation based on projected returns. *Neuropsychologia* **1492**, 1–8 (2002).
17. di Bitetti, M. S. Evidence for an important social role of allogrooming in a platyrrhine primate. *Anim. Behav.* **54**, 199–211 (1997).
18. de Waal, F. B. M. Food transfers through mesh in brown capuchins. *J. Comp. Psychol.* **111**, 370–378 (1997).
19. Tinklepaugh, O. L. An experimental study of representative factors in monkeys. *J. Comp. Psychol.* **8**, 197–236 (1928).
20. Andreoni, J., Brown, P. M. & Vesterlund, L. What makes an allocation fair? Some experimental evidence. *Games Econ. Behav.* **40**, 1–24 (2002).
21. Falk, A., Fehr, E. & Fischbacher, U. On the nature of fair behavior. *Econ. Inq.* **41**, 20–26 (2003).
22. Frank, R. H. in *Evolution and the Capacity for Commitment* (ed. Nesse, R. M.) 57–76 (Russell Sage Foundation, New York, 2001).
23. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural basis of economic decision-making in the Ultimatum game. *Science* **300**, 1755–1758 (2003).
24. Frank, R. H. *Passions Within Reason: The Strategic Role of the Emotions* (W. W. Norton, New York, 1988).
25. Hirschleifer, J. in *The Latest on the Best: Essays in Evolution and Optimality* (ed. Dupre, J.) 307–326 (MIT Press, Cambridge, Massachusetts, 1987).
26. Nesse, R. M. in *Evolution and the Capacity for Commitment* (ed. Nesse, R. M.) 1–44 (Russell Sage Foundation, New York, 2001).
27. Dugatkin, L. A. in *Evolution and the Capacity for Commitment* (ed. Nesse, R. M.) 120–137 (Russell Sage Foundation, New York, 2001).
28. de Waal, F. B. M. & Berger, M. L. Payment for labour in monkeys. *Nature* **404**, 563 (2000).
29. Mendres, K. A. & de Waal, F. B. M. Capuchins do cooperate: the advantage of an intuitive task. *Anim. Behav.* **60**, 523–529 (2000).
30. Brosnan, S. F. & de Waal, F. B. M. A proximate perspective on reciprocal altruism. *Hum. Nat.* **13**, 129–152 (2002).

Acknowledgements We thank L. Antonucci, K. Boussein, J. Davis, C. Freeman and A. Katz for help with data collection and coding, L. Rutten for comments on the manuscript and R. Earley for statistical advice. Photo by E.D.W.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.F.B. (sbrosna@emory.edu).

cdx4 mutants fail to specify blood progenitors and can be rescued by multiple hox genes

Alan J. Davidson¹, Patricia Ernst², Yuan Wang³, Marcus P. S. Dekens⁴, Paul D. Kingsley⁵, James Palis⁵, Stanley J. Korsmeyer², George Q. Daley³ & Leonard I. Zon¹

¹Department of Medicine, Division of Hematology/Oncology, Children's Hospital and Dana-Farber Cancer Institute, Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

²Departments of Pathology and Medicine, Dana-Farber Cancer Institute and Harvard Medical School, Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

³Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

⁴Max-Planck-Institut für Entwicklungsbiologie, Abteilung Genetik, Spemannstrasse 35, 72076 Tübingen, Germany

⁵Department of Pediatrics, Center for Human Genetics and Molecular Pediatric Diseases, University of Rochester, Rochester, New York 14642, USA

Organogenesis is dependent on the formation of distinct cell types within the embryo. Important to this process are the *hox* genes, which are believed to confer positional identities to cells along the anteroposterior axis^{1–3}. Here, we have identified the *caudal*-related gene *cdx4* as the locus mutated in *kugelig* (*kgg*), a zebrafish mutant with an early defect in haematopoiesis that is associated with abnormal anteroposterior patterning and aberrant *hox* gene expression. The blood deficiency in *kgg* embryos can be rescued by overexpressing *hoxb7a* or *hoxa9a* but not *hoxb8a*, indicating that the haematopoietic defect results from perturbations in specific *hox* genes. Furthermore, the haematopoietic defect in *kgg* mutants is not rescued by *scl* overexpression, suggesting that *cdx4* and *hox* genes act to make the posterior mesoderm competent for blood development. Overexpression of *cdx4* during zebrafish development or in mouse embryonic stem cells induces blood formation and alters *hox* gene expression. Taken together, these findings demonstrate that *cdx4* regulates *hox* genes and is necessary for the specification of haematopoietic cell fate during vertebrate embryogenesis.

The yolk sac blood islands of amniotes develop from posterior mesoderm and form embryonic erythroid cells and endothelial cells. The equivalent site in zebrafish, known as the intermediate cell mass (ICM), arises from bilateral stripes of haematopoietic and vascular precursors in the posterior mesoderm. We found that embryos homozygous for *kgg*, an autosomal recessive mutation that was initially identified because of tail defects⁴, exhibit severe anaemia within the first day of development. Although blood cell numbers begin to recover by 5 days post-fertilization (d.p.f.), all mutants die between 7 and 10 d.p.f.

To investigate the haematopoietic defect in *kgg*, we examined the expression of *scl*, *gata1* and *runx1* genes. At the 5-somite stage, the bilateral stripes of *scl*⁺ cells are thinner in *kgg*^{tv205} embryos than in wild-type controls (Fig. 1b). In addition, *kgg*^{tv205} mutants show a decreased number of *gata1*⁺ erythroid precursors and a complete absence of *runx1* expression in blood and neuronal cells. By 24 h post-fertilization (h.p.f.), *kgg*^{tv205} mutants have a severe reduction in the number of haemoglobin-expressing erythroid cells compared with wild-type siblings (Fig. 1b). By contrast, normal numbers of *pu.1*⁺ myeloid cells are formed from the cephalic mesoderm in *kgg*^{tv205} embryos (data not shown). To study the development of the vasculature in the mutant, we examined the expression of the VEGF receptor *flk1*. At the 10- and 15-somite stages, *kgg*^{tv205} embryos have relatively normal numbers of angioblasts, although their convergence to the midline is delayed (Fig. 1c). By 24 h.p.f., the vasculature

appears well formed in the mutants and the few blood cells that develop circulate normally. The pronephric kidney arises from mesoderm adjacent to the ICM precursors. In *kgg*^{tv205} mutants, the expression domains of the pronephric duct markers *pax2.1* and *cxc4b* are shortened (arrowheads in Fig. 1d), although unlike the *scl* stripes, the width of the *pax2.1* stripe is unaffected. Transcripts for the glomerulus marker *wt1*, which are normally expressed in mesoderm adjacent to somites one to four, extend from somites one to six in *kgg*^{tv205} embryos (brackets in Fig. 1d), suggesting that the *kgg*^{tv205} mutation leads to an expansion of anterior kidney fates at the expense of more posterior fates. Other structures such as the head, notochord and somites appear grossly normal in *kgg*^{tv205} embryos, although the length of the embryo is shortened compared with wild-type embryos.

The *kgg* mutation maps to linkage group 14 (Fig. 2a) near candidate genes including *cdx4*, *smad5* and *wnt8*. An analysis of the complementary DNA sequence of *wnt8* and *smad5* from *kgg* mutants did not identify any mutations. *cdx4* belongs to the *caudal* family of homeobox genes, which have been implicated in anteroposterior patterning^{5–7}. Three *caudal* paralogues exist in mammals (*Cdx1*, *Cdx2* and *Cdx4*), and mouse gene-targeting studies of *Cdx1* and *Cdx2* (*Cdx4* has yet to be targeted) have demonstrated a role for these genes in the anteroposterior patterning of the axial skeleton^{8–10}. In addition, *Cdx2*^{+/-} mice develop hamartomatous polyps in the colon that result from a transformation of the intestinal epithelium to a more anterior (gastric) fate^{9,11,12}. Sequence analysis of the *cdx4* gene from *kgg*^{tv205} mutants revealed a T to A transversion in nucleotide +510, changing a conserved F170 residue in the homeodomain to a leucine (Fig. 2b). This mutation prevents the protein from binding to a *Cdx4* consensus binding site in gel-shift experiments (see Supplementary Fig. 1). A partial deletion of the *cdx4* gene, and at least one other neighbouring gene (*chic1*), was found in *kgg*^{tv205} mutants (Fig. 2a, b; see also Supplementary Fig. 2). We isolated the *cdx4* transcript in *kgg*^{tv205} mutants by 3' rapid amplification of cloned ends (RACE) and found that exon two had become spliced onto downstream sequence that extended the *cdx4* open reading frame by 11 amino acids (GFSSVFQSQSD-stop). Radiation hybrid mapping of this foreign sequence placed it 20 cR away from the *cdx4* locus. To provide further evidence that the *kgg* mutant phenotype is caused by defects in *cdx4*, we injected wild-type embryos with *cdx4* antisense morpholino oligonucleotides and found that the resulting morpholino mutants (known as morphants) phenotypically resembled *kgg* embryos (Figs 1b and 2c).

Transcripts for *cdx4* are first detected in the early gastrula but become restricted to the posterior-most cells during gastrulation and early somitogenesis (Fig. 2d). Double whole-mount *in situ* hybridization and sectioning at the 3-somite stage revealed that the *cdx4* expression domain initially includes cells in the posterior mesoderm that express *scl* (Fig. 2d and data not shown). However, from the 5-somite stage onwards the expression domains of *cdx4* and *scl* are largely non-overlapping. Similar expression profiles were found for the mouse orthologues of *cdx4* and *scl* during early embryogenesis (Fig. 2e). At the late primitive streak stage (E7.25), *Cdx4* transcripts are confined to mesodermal cells of the posterior embryo, the allantois and the forming yolk sac wall. Although *Cdx4* is not expressed in the nascent blood islands, its expression domain does partially overlap with *Scl* in mesodermal cells of the posterior primitive streak and the posterior yolk sac.

To further explore the function of *cdx4* during embryonic haematopoiesis, we examined the effect of *cdx4* overexpression in wild-type embryos. Embryos injected with *cdx4* messenger RNA (7, 15 or 30 pg) display a range of 'posteriorized' phenotypes (see Supplementary Fig. 3a, b). By contrast, embryos injected with 15 pg of F170L mutant mRNA all exhibit a wild-type morphology (*n* = 60 of 60 embryos injected; data not shown). The effect of *cdx4* overexpression (15 pg) on blood development was examined at the 5- to 12-somite stages. Surprisingly, 12–20% of the injected embryos

showed ectopic *scl* ($n = 24$ of 118), *gata1* ($n = 7$ of 59) and *fli1* ($n = 4$ of 26) expression near the midline in a stripe that ran parallel to the endogenous blood precursors (Fig. 3a–c). Cross-sections revealed that the ectopic *scl*⁺ cells were unilaterally located adjacent to the notochord (Fig. 3b). The reason for this restricted localization is currently unclear; however, the genes induced appear to be specific to the haematopoietic programme, as ectopic *flk1* expression was confined to the upper trunk region ($n = 11$ of 69; Fig. 3c), whereas no ectopic expression of *pax2.1* was found ($n = 0$ of 55; data not shown). By contrast, 11–22% of the injected embryos exhibited decreased expression of *scl*, *gata1*, *fli1*, *flk1* and *pax2.1* (data not shown). The disrupted tissue development in these embryos may result from abnormal gastrulation, or the conversion of mesoderm to an extreme posterior fate. To assess the ability of *cdx4* to rescue *kgg*^{tv205} mutants, we injected 15 pg of *cdx4* mRNA and assayed the number of *scl*⁺ and *gata1*⁺ cells at the 5- and 10-somite stages, respectively. Consistent with *cdx4* being the gene

defective in *kgg* mutants, the haematopoietic defects were partially rescued in approximately 80% of injected mutants ($n = 15$ of 19 mutants for *scl* and $n = 27$ of 33 mutants for *gata1*; Fig. 3d, e).

caudal homologues have been implicated in anteroposterior patterning by regulating the expression of *hox* genes^{8,13,14}. To investigate *hox* gene expression in *kgg* mutants, we examined the expression of selected *hoxb* cluster genes and *hoxa9a*, as many of these genes are known to affect haematopoiesis¹⁵. All of the *hox* genes examined (*hoxb4*, *hoxb5a*, *hoxb6b*, *hoxb7a*, *hoxb8a*, *hoxb8b* and *hoxa9a*) display altered expression patterns in *kgg*^{tv205} embryos (Fig. 4a). For instance, the mesodermal expression of *hoxb5a* normally includes somites two and three, the notochord and the tailbud region, but in *kgg*^{tv205} mutants, *hoxb5a* expression is expanded to include somites two to five, is absent from the notochord and is reduced in the tailbud (Fig. 4a). In the case of *hoxb6b* and *hoxa9a*, the expression of these *hox* genes is almost absent in *kgg*^{tv205} mutants.

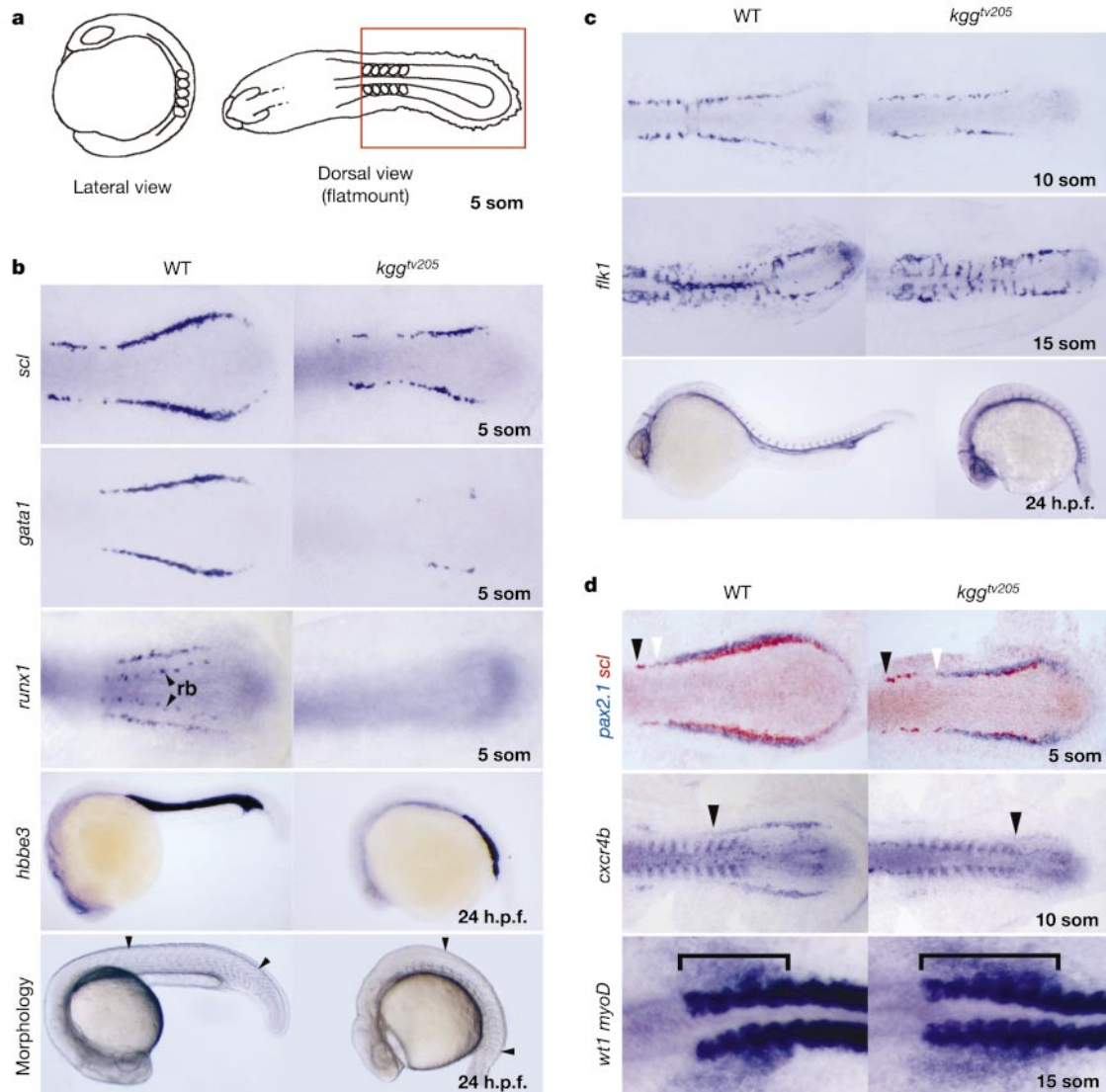


Figure 1 Abnormal blood and kidney development in *kgg*^{tv205} mutants. **a**, Schematic representation of 5-somite-stage embryos. Boxed region indicates views in **b–d**. **b**, Expression of *scl*, *gata1*, *runx1* (arrowheads indicate the stripes of Rohon–Beard neurons) and *haemoglobin beta embryonic3* (*hbbe3*) in wild-type (WT) and *kgg*^{tv205} embryos. Live morphologies of wild-type and *kgg*^{tv205} embryos are shown in the bottom panels. In wild-type embryos, the ICM extends from the level of somite 5 to 18 (indicated

with arrowheads). **c**, Expression of *flk1* in wild-type and *kgg*^{tv205} embryos. **d**, Expression of *pax2.1* (purple) and *scl* (red) in wild-type and *kgg*^{tv205} embryos (upper panels). Black and white arrowheads indicate the anterior limits of *scl* and *pax2.1* expression, respectively. Middle panels show expression of *cxcr4b* in wild-type and *kgg*^{tv205} embryos. Arrowheads indicate the anterior limit of pronephric duct expression. Lower panels show expression of *wt1* (bracket) and the somite marker *myoD* in wild-type and *kgg*^{tv205} embryos.

To further understand how the stripe of haematopoietic/vascular precursors is affected by changes in anteroposterior patterning, we examined the *scl*⁺ populations in more detail. During normal development, transcripts for *scl* are first detected around the 3-somite stage in stripes of mesoderm adjacent to the future site of somite six (arrowheads in Fig. 4b). At the 5-somite stage, *de novo* expression of *scl* occurs adjacent to somites one to five (bracket in Fig. 4b). These cells are probably angioblasts as they express *flk1* but not *gata1* (Fig. 4c). Transcripts for *flk1* and *gata1* in cells of the posterior *scl*⁺ stripe appear mutually exclusive, suggesting that this stripe comprises juxtaposed populations of angioblasts and haematopoietic precursors.

In *kgg* mutants, there is a preferential loss of *gata1*⁺ haematopoietic cells from the posterior stripe with little effect on the adjacent angioblasts. This blood loss in *kgg* mutants may result, in part, from a posterior shift in the boundary between the anterior (angioblast) and posterior (blood and angioblast) *scl*⁺ populations. In support of this, the expression domains of *hoxb6b*, *hoxb7a* and *hoxa9a*, which share an anterior expression limit with *gata1* (Fig. 4c and data not shown), are significantly reduced in *kgg*^{iv205} mutants as early as the 3-somite stage (Fig. 4d). By contrast, the *scl*⁺ anterior angioblasts are found rostral to the *hoxb7a* expression domain but at a similar anteroposterior level to *hoxb5a* (Fig. 4c). Given that *hox* gene overexpression can transform cell fates¹⁶, we examined

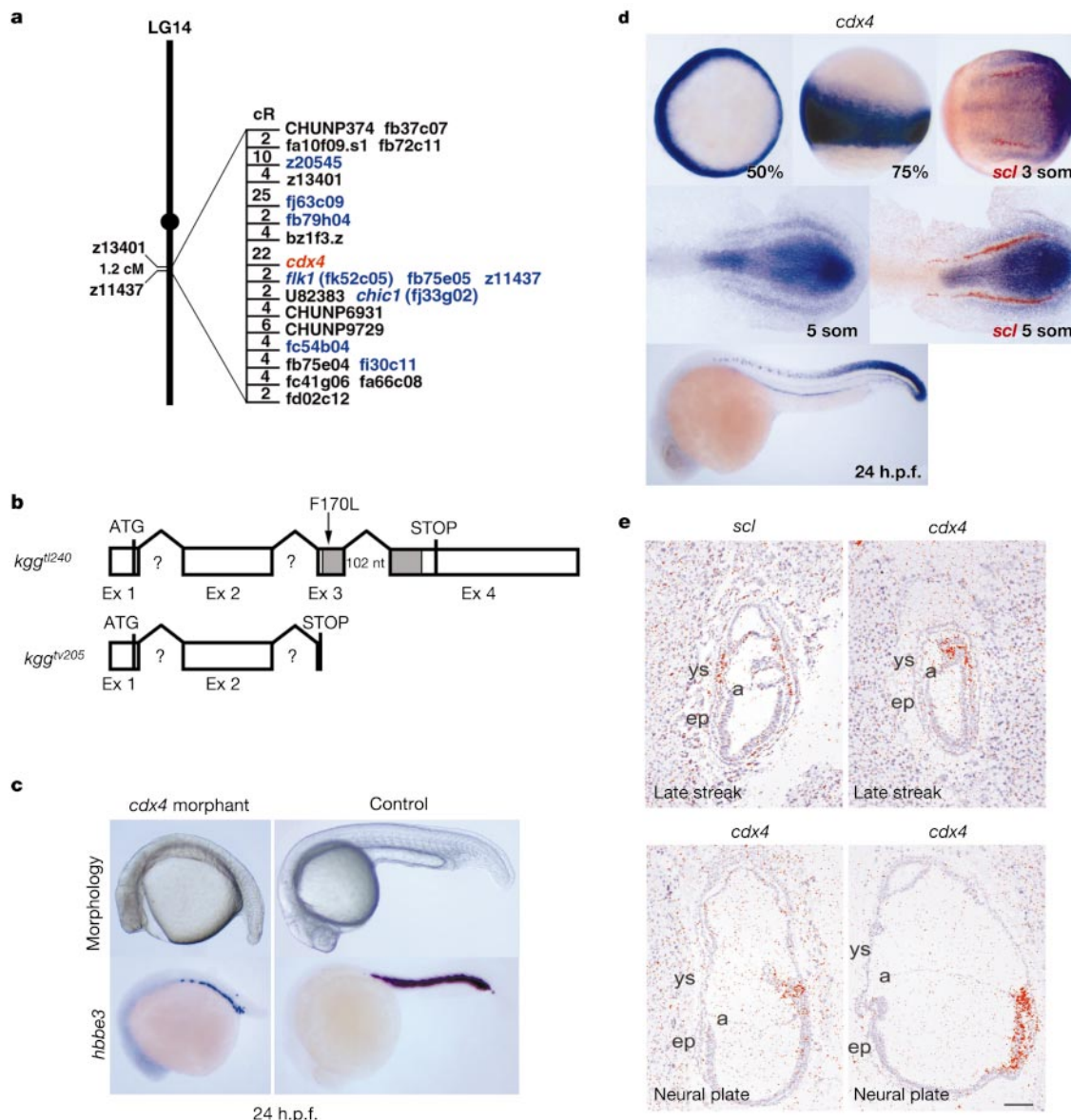


Figure 2 Isolation of the *kgg* gene and developmental expression of *cdx4*. **a**, Schematic representation of linkage group (LG) 14 and the corresponding region of the Goodfellow radiation hybrid panel. *cdx4* is shown in red and expressed sequence tags/markers tested for deletion in the *kgg*^{iv205} allele are blue (see Supplementary Fig. 2). **b**, Schematic representation of the *cdx4* gene and the defects found in the *kgg*^{iv240} and *kgg*^{iv205} alleles. The regions of exon three and four that encode the homeodomain are shaded and the intervening intron of 102 nucleotides (nt) and the positions of the F170L mutation (*kgg*^{iv240}) and deletion (*kgg*^{iv205}) are indicated. In the *kgg*^{iv205} mutant, deletion of exons three and four leads to abnormal splicing of exon two to downstream sequence (black exon). **c**, Phenocopy of the *kgg* mutation by morpholino oligonucleotide-mediated knock

down of *cdx4*. The morphology and expression of *hbbe3* in a *cdx4* morphant and control embryo at 24 h.p.f. is shown. **d**, Developmental expression of *cdx4* (purple) at the 50% epiboly (animal pole view, dorsal to right), 75% epiboly (lateral view, dorsal to right), 3-somite (dorso-posterior view, anterior to left), 5-somite (flat-mounted with anterior to left) and 24 h.p.f. (lateral view, anterior to left) stages. Embryos at the 3- and 5-somite stages were double-stained for both *cdx4* (purple) and *scl* (red) transcripts. **e**, Expression of murine *Cdx4* and *Scl* during gastrulation. Anterior is to the left and posterior is to the right in each panel. Signal indicated by silver grains is pseudocoloured red. The amnion (a) marks the separation of the yolk sac (ys) from the embryo proper (ep). Scale bar, 100 μ m.

whether overexpression of *hox* paralogues from the 6th, 7th, 8th or 9th groups were capable of rescuing the blood defect in *kkg^{tv205}* mutants. Mutants injected with 3 pg of *hoxb7a* and *hoxa9a* mRNA displayed an almost complete rescue of *gata1*⁺ blood cells at the 18-somite stage (65%, *n* = 13 of 20 mutants and 100%, *n* = 18 of 18, respectively; Fig. 4e), although the axial and tail defects were not rescued. By contrast, the highest non-toxic dose of *hoxb6b* mRNA (1–2 pg; 64%, *n* = 7 of 11) led to a small increase in *gata1*⁺ blood cells, whereas the highest non-toxic level of *hoxb8a* mRNA (1–2 pg) failed to rescue the blood defects (*n* = 0 of 22 mutants; data not shown). Taken together, these findings suggest that the specification of haematopoietic cell fate is dependent on the proper expression of *hox* genes such as *hoxb7a* and *hoxa9a* in the posterior mesoderm, and that overexpression of any one of these *cdx4* targets can rescue erythropoiesis in *kkg* mutants.

To provide further evidence that *cdx4* and *hox* genes function together in a common pathway, we examined whether *cdx4* overexpression (15 pg) could rescue the expression of *hoxb6b*, *hoxb7a* and *hoxa9a* in *cdx4* morphants. We found a restoration of *hoxb6b*, *hoxb7a* and *hoxa9a* expression domains in *cdx4*-rescued morphants (Fig. 5a). Interestingly, approximately 80% of the injected embryos also displayed ectopic *hoxb7a* expression in the forebrain and/or hindbrain regions (*n* = 31 of 39; arrowheads in Fig. 5a), supporting a role for *cdx4* in the induction of *hox* gene expression.

In zebrafish, overexpression of *scl* leads to an expansion of haematopoietic cells in the posterior lateral plate mesoderm¹⁷. We examined whether *scl* overexpression could rescue erythropoiesis in *kkg* mutants. Wild-type embryos injected with *scl* mRNA (100 pg)

display an expanded number of *gata1*⁺ erythroid precursors at the 10-somite stage (Fig. 5b). By contrast, no such expansion in erythroid cell numbers was found in *scl*-injected *kkg* embryos (Fig. 5b). Given that *cdx4* expression precedes that of *scl* in the posterior mesoderm, our results suggest that the specification of haematopoietic fate by *scl* is dependent on *cdx4*.

Several studies have shown that retroviral expression of *HoxB4* in haematopoietic stem cells or multipotential progenitors enhances their self-renewal/proliferation^{18–20}. To examine whether *Cdx4* has a similar activity, we retrovirally transduced murine embryoid body haematopoietic cells with *Cdx4* and assayed the effect on multilineage haematopoietic colony formation. In this system, *Cdx4* induced a pronounced expansion of haematopoietic progenitors, including a 13-fold increase in CFU-GEMM (colony forming unit-granulocyte/erythroid/macrophage/megakaryocyte) colonies and an 11-fold increase in CFU-GM colonies compared with control cells (Fig. 5c). The *Cdx4*-mediated expansion of multilineage progenitors and colony size was more potent than that observed with *HoxB4*, which induced a 9-fold increase in CFU-GEMM (Fig. 5c). We next examined changes in the expression of selected *HoxA*, *HoxB* and *HoxC* cluster genes in the *Cdx4*-transduced cells, using quantitative polymerase chain reaction (PCR) (Fig. 5d). *Cdx4* induced widespread alterations in *Hox* expression levels in transduced cells including a marked increase in the expression of *HoxB4* (30-fold), *HoxB3* (19-fold), *HoxB8* (5-fold) and *HoxA9* (4.1-fold), all of which have been implicated in haematopoietic stem cell or immature progenitor expansion^{21–23}.

During embryoid body development, *Cdx4* is expressed at days 3

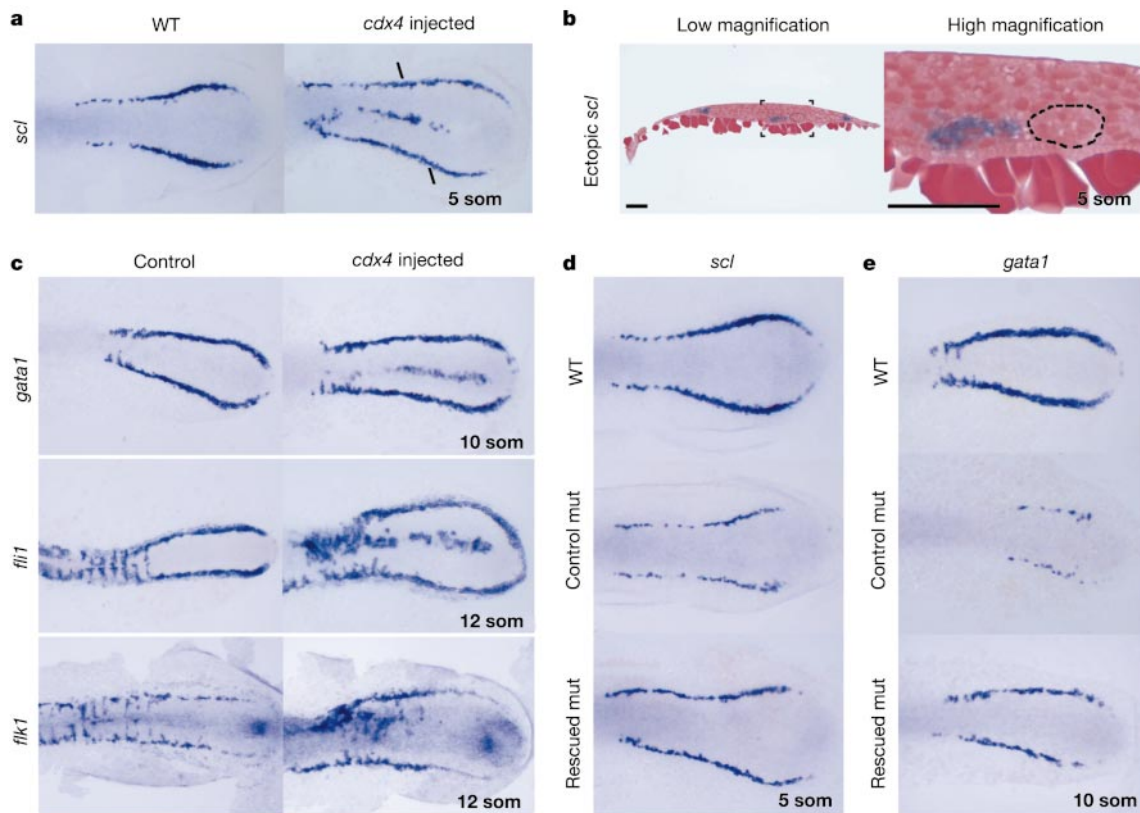


Figure 3 Overexpression of *cdx4* induces ectopic blood and rescues *kkg^{tv205}* embryos. **a**, Expression of *scl* in an uninjected (wild type) or a *cdx4*-injected embryo. The black lines in the second panel indicate the level of the cross-section shown in **b**. **b**, Cross-section of the embryo in **a** showing ectopic *scl*-expressing cells (purple) near the notochord (outlined). A magnified view of the indicated area is shown in the adjacent panel. Scale bars, 50 μ m. **c**, Expression of *gata1*, *flil1* and *flk1* in uninjected (control) or *cdx4*-injected

embryos. The *flil1* gene is a marker of haematopoietic precursors and angioblasts. **d**, **e**, Partial rescue of *scl*- and *gata1*-expressing cells in *kkg^{tv205}* mutants by injection of *cdx4* mRNA. Panels show expression of *scl* and *gata1* in uninjected wild-type embryos (top panels), uninjected *kkg^{tv205}* embryos (middle panels) and in rescued *kkg^{tv205}* mutants injected with 15 pg of *cdx4* mRNA (bottom panels).

and 4 (Fig. 5e), during the time when haematopoietic precursors first arise. To narrow down the time window during embryoid body differentiation in which *Cdx4* enhances multilineage haematopoietic colony formation, we engineered embryonic stem cells to express *Cdx4* under the control of a tetracycline-inducible promoter (data not shown). The strongest effect of *Cdx4* on colony formation was found between days 4 and 5 of embryoid body development, with increased multipotent progenitors (CFU-GEMM), CFU-GM and primitive erythroid colonies compared with uninduced embry-

oid bodies (Supplementary Fig. 4). Taken together, these results are consistent with *Cdx4* acting at early stages of haematopoietic development to enhance the proliferation of early haematopoietic progenitors by upregulating the expression of target *Hox* genes.

Our analysis of *kkg* supports a role for *hox* genes in the development of the haematopoietic system. Loss of *cdx4* activity causes widespread perturbations in *hox* expression domains that are manifested as a posterior shift in the boundary between the anterior endothelial population and the more posterior populations of

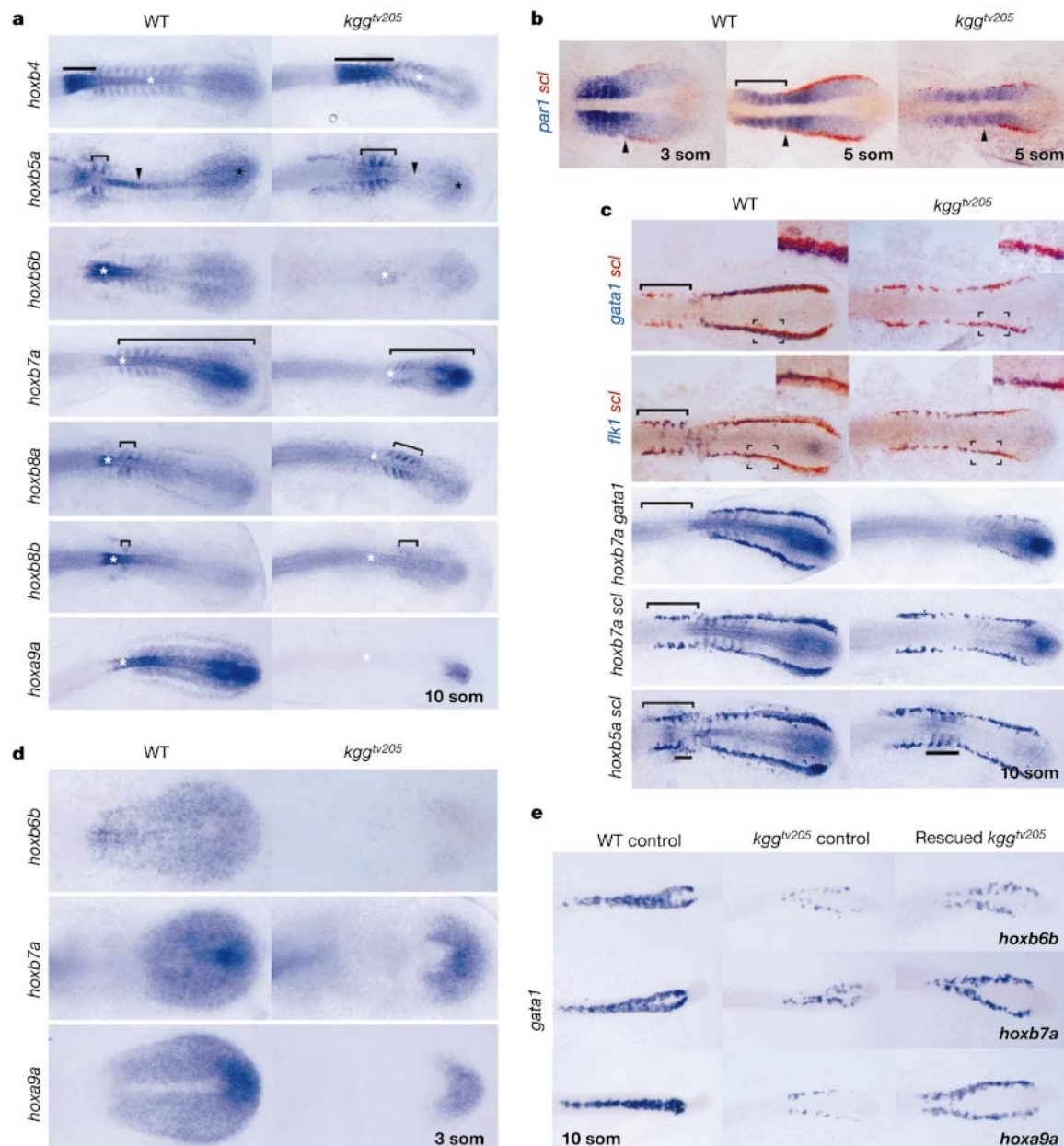


Figure 4 The haematopoietic defect in *kkg*^{tv205} mutants is associated with abnormal *hox* gene expression and can be rescued by *hoxb6b*, *hoxb7a* and *hoxa9a*. **a**, Expression patterns of *hoxb4*, *hoxb5a*, *hoxb6b*, *hoxb7a*, *hoxb8a*, *hoxb8b* and *hoxa9a* in the central nervous system (black bars and white asterisks), paraxial mesoderm (brackets), notochord (arrowheads) and tailbud region (black asterisks) of wild-type and *kkg*^{tv205} embryos. **b**, Expression of *scl* (red) and *paraxis1* (*par1*; purple) in wild-type and *kkg*^{tv205} mutants. The *par1* gene is expressed in the paraxial mesoderm during somite formation. The arrowhead indicates the level of somite six and the bracket demarcates the anterior population of putative angioblasts. **c**, Expression of haematopoietic (*scl*, *gata1*) and

vascular (*scl*, *flk1*) markers relative to *hoxb7a* and *hoxb5a* in wild-type and *kkg*^{tv205} embryos. Insets show higher magnification of the indicated areas, the bracket indicates the anterior population of angioblasts, and black bars demarcate somitic expression of *hoxb5a*. For the lower three panels both genes are stained in purple. **d**, Expression of *hoxb6b*, *hoxb7a* and *hoxa9a* at the 3-somite stage in wild-type and *kkg*^{tv205} mutants. Note the severe reduction in *hoxb6b* and *hoxa9a* expression. **e**, Expression of *gata1* in wild-type embryos, *kkg*^{tv205} mutants and rescued *kkg*^{tv205} embryos injected with *hoxb6b*, *hoxb7a* or *hoxa9a* mRNA as indicated.

blood and endothelial cells. In addition, there is an overall reduction in erythroid cell numbers (schematically represented in Fig. 5f). These findings also have implications for the concept of the haemangioblast, a putative bipotential cell that is thought to express *scl* and give rise to both blood and vascular lineages *in vivo*. *kkg* mutants display a reduced number of *scl*⁺ cells with a selective loss of blood but not angioblasts. This result suggests that if haemangioblasts exist *in vivo* then they must arise before the onset of *scl* expression, and that *cdx4* is necessary for this population to differentiate into an *scl*⁺ haematopoietic precursor. Alternatively, the blood and vascular lineages may arise independently from the posterior mesoderm, with *cdx4* being required solely for the specification of haematopoietic fate. The finding that *scl* overexpression fails to rescue blood development in *kkg* mutants

suggests that the *cdx4*–*hox* pathway acts to make the posterior lateral plate mesoderm competent to respond to genes that specify haematopoietic fate.

It is likely that multiple *hox* genes with redundant activities participate in blood development. In support of this, the targeted disruption of *HoxB6*, *HoxB7* or *HoxA9* in mice does not block early embryonic haematopoiesis^{24–26}. Similarly, we have been unable to find single or combinations of *hox* gene morpholino oligonucleotides that inhibit blood formation during zebrafish development, although nonspecific toxicity makes it difficult to inject more than three morpholino oligonucleotides simultaneously. Deregulated expression of *Hox* genes has been implicated in leukaemic transformation¹⁵. The function of *cdx* genes as transcriptional regulators of *hox* genes raises the possibility that this family may also

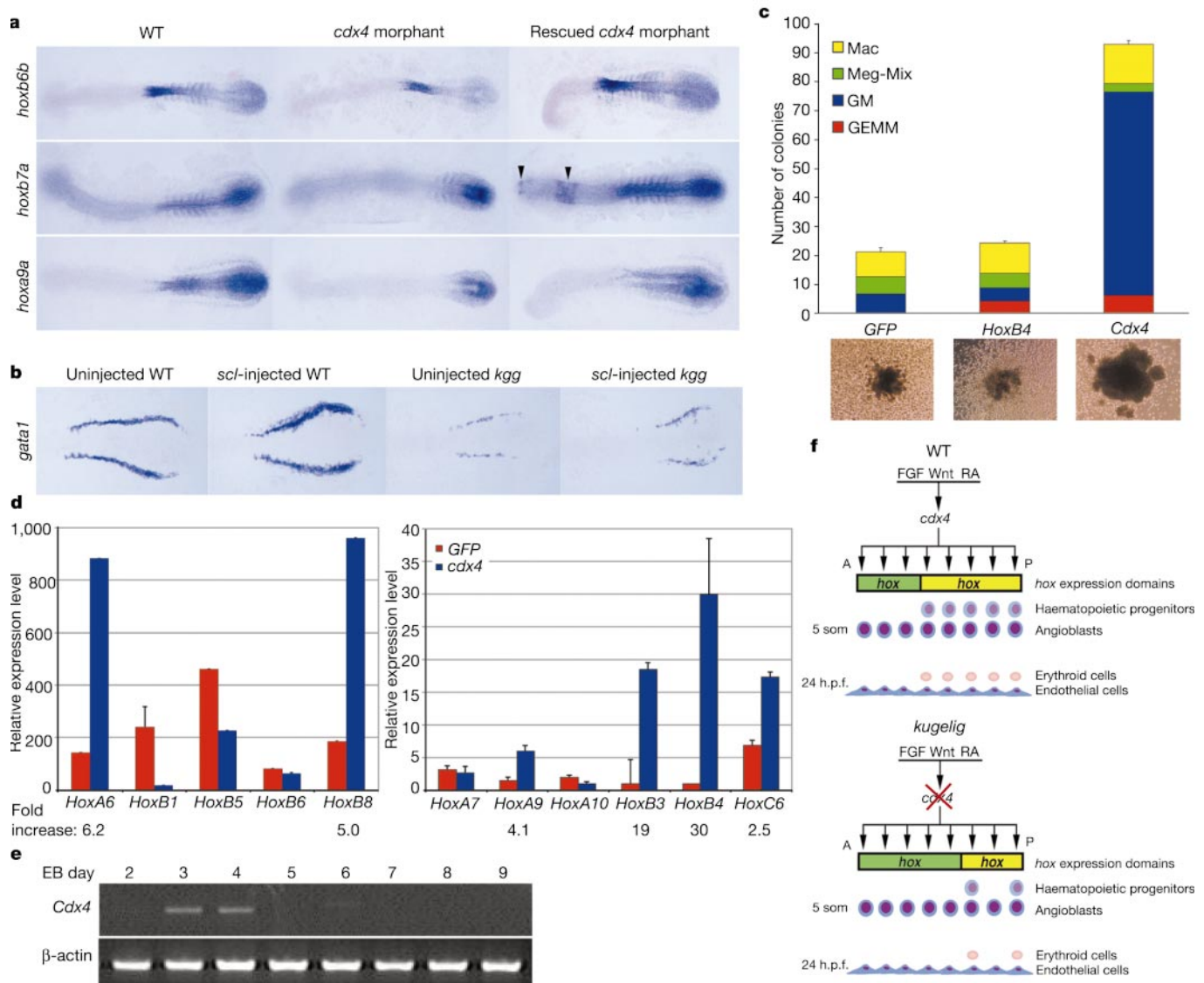


Figure 5 *Cdx4* alters *hox* gene expression in zebrafish and mouse cells and induces blood development in embryoid bodies. **a**, Expression of *hoxb6b*, *hoxb7a* and *hoxa9a* at the 10-somite stage in wild-type embryos, *cdx4* morphants and *cdx4* morphants injected with *cdx4* mRNA. Arrowheads indicate ectopic *hoxb7a* expression in the forebrain and hindbrain. **b**, Expression of *gata1* at the 10-somite stage in an *scl*-injected wild-type and *kkg*^{h205} mutant and in uninjected controls. **c**, Effect of *Cdx4* and *HoxB4* overexpression on haematopoietic progenitors derived from embryoid bodies. Colony forming units scored are macrophage (Mac), megakaryocytes and mixed lineage (Meg-mix), granulocyte/macrophage (GM), and granulocyte/macrophage/megakaryocyte (GEMM).

Bottom panels show images of representative colonies. **d**, Quantitative PCR analysis of the expression of selected *HoxA*, *HoxB* and *HoxC* cluster genes in embryoid bodies overexpressing *Cdx4*. **e**, RT–PCR analysis of *Cdx4* expression during embryoid body (EB) development. **f**, Model for the role of *cdx4* in anteroposterior patterning and blood development. Signalling molecules such as fibroblast growth factors (FGFs), Wnts and retinoic acid (RA) are known to regulate the expression of *cdx4*, which in turn establishes the correct expression domains of *hox* genes necessary for blood development. In the absence of *cdx4* (bottom panel), *hox* expression domains are shifted and fewer erythroid cells are formed.

participate in leukaemogenesis. Consistent with this, a fusion of *CDX2* to *TEL/ETV6*, a gene frequently rearranged in haematological malignancies, has been found in a patient with acute myeloid leukaemia²⁷. The challenge for future studies will be to better understand how the *cdx4*–*hox* pathway regulates commitment to a haematopoietic fate and participates in leukaemia. □

Methods

Deletion analysis and genotyping

The extent of the *kcg*^{tr205} deletion was determined by PCR using the following primers to *cdx4*: exon one (forward 5′-AGCTCCTTTTGGACTATTAC-3′; reverse 5′-CCAACGTA CATGATTGGAA-3′), exon two (forward 5′-ATACCTTTTGGAGAAAGAGG-3′; reverse 5′-CCGGTTGATGACGACTGGAC-3′), exon three (forward 5′-CAAAACGAGAACGAA GGAGA-3′; reverse 5′-ACCTGTCTCTCTGAAAGCCC-3′) and exon four (forward 5′-TAAGATCTGGTTTCAGAAC-3′; reverse 5′-TGGATGATCCAAGTTCGAGT-3′). Exon three forward and exon four reverse primers were used to genotype *kcg*^{tr205} embryos. Primers to expressed sequence tags mapping near *cdx4* were obtained from the WashU Zebrafish Genome Resources Project (<http://zfsh.wustl.edu/>). Primers for z20545 and z11437 were obtained from the Massachusetts General Hospital Zebrafish Server (<http://zebrafish.mgh.harvard.edu/>).

Electrophoretic mobility shift assays

Double-stranded oligonucleotide probes contained a single consensus Cdx-binding site (5′-GAGAAATTTATATTGT-3′; consensus italicized) or mutated site (5′-GAGAAAT CCATATTGT-3′; mutated nucleotides italicized). ³⁵S-methionine-labelled Cdx4 (wild type) and the F170L mutant proteins were resolved on a 10–20% Tris-HCl polyacrylamide gel (Ready Gels, Biorad).

Inducible *Cdx4* embryonic stem cell lines and colony assays

Cdx4 was subcloned into the plox vector²⁸ and electroporated into Ainv15 embryonic stem cells together with pSalk-Cre, followed by selection with G418 (400 µg ml^{−1}). Colonies positive for plox-*Cdx4* were confirmed by RT-PCR. The tetracycline-inducible *Cdx4* embryonic stem cells and embryoid bodies were maintained and produced as described previously²⁸. Embryoid bodies were collected at day 6 by collagenase treatment and plated into methylcellulose (M3434, StemCell Technologies). Colonies were scored 6–9 days later.

Microinjection

Wild-type and F170L mutant *cdx4* cDNAs were subcloned into pCS2⁺ for mRNA synthesis. Messenger RNAs were injected between the 1–4 cell stages at a concentration of 30 ng µl^{−1}. Full-length *hox* genes were amplified from 5-somite-stage cDNA by RT-PCR and subcloned into pCS2⁺. Messenger RNAs were injected at a concentration of 200, 6 and 2–4 ng µl^{−1}, for *scl*, *hoxb7/hoxa9a* and *hoxb6b/hoxb8a*, respectively. The *cdx4* morpholino oligonucleotides (CGTACATGATTGGAAGAAACCCCT; start codon italicized) were obtained from Gene Tools LLC and solubilized in X1 Danieau solution. Embryos were injected with 1 nl *cdx4* morpholino oligonucleotide at a concentration of 0.2 mg ml^{−1}.

Mutation analysis by RT-PCR

Complementary DNA was prepared from *kcg* mutants and wild-type embryos at 24 h.p.f. The *cdx4* open reading frame was amplified using forward (5′-CATGTACGTGGATACC TTTTGG-3′) and reverse (5′-TCCACAACCCACGCTCTTATT-3′) primers, subcloned into pGEM-T and sequenced. Our cDNA sequence of *cdx4* differs from the published sequence (GenBank NM_131109) by the addition of two cytosine nucleotides at +709–710. The resulting frameshift changes the open reading frame of the carboxy terminus to give a predicted protein of 271 residues rather than the published length of 301 residues²⁹.

Radiation hybrid and genetic mapping

The genetic map position of *kcg* was obtained from the Max-Planck-Institut für Entwicklungsbiologie (<http://www.map.tuebingen.mpg.de/>). The *cdx4* gene was mapped onto the Goodfellow RH panel by the Children's Hospital Genome Initiative group using forward (5′-AGGCGTGGGTGTGGATTAC-3′) and reverse (5′-GATACACTACCAC ATACAG-3′) primers. The genomic contig encoding the foreign exon in *kcg*^{tr205} mutants was mapped using forward (5′-GTGATCAACAACACGTCC-3′) and reverse primers (5′-GGAATCTCCTGTCAGCTG-3′).

Cdx4 retroviral expression and quantitative PCR

Murine *Cdx4* was subcloned into pMSCV-IRES-GFP and retroviruses were generated using an ecotropic packaging vector and co-transfection. Embryoid bodies were formed from RW4 embryonic stem cells by differentiating for 6 days and definitive haematopoietic cells were enriched tenfold using an anti-CD41 magnetic strategy. Approximately 1,000,000 cells were plated on OP9 monolayers and subjected to two rounds of retroviral infection with either GFP only or *Cdx4*–GFP retroviruses. After 48 h, GFP⁺ cells were sorted and either lysed in Trizol (Invitrogen), or plated in methylcellulose (M3434, StemCell Technologies) and scored for colony types 3–7 days later. Complementary DNA was prepared from GFP-expressing or *Cdx4*–GFP-expressing cells and real-time PCR was performed with an ABI Prism 7700 Sequence Detector and dual-labelled probes (sequence available on request), with the exception of *HoxB4*, which was quantified using Sybr green (Applied Biosystems). *GAPDH* was used to normalize samples. *Hox* expression levels are expressed in arbitrary units (relative to the lowest sample) using the comparative C_T method.

In situ hybridization and sectioning

In situ hybridization of zebrafish and mouse embryos was performed as previously described^{17,30}. Zebrafish embryos to be sectioned were infiltrated in JB-4 resin, cut at a thickness of 5 µm, and counterstained in 0.5% safranin O. Sections of mouse embryos were counterstained with haematoxylin.

Received 27 June; accepted 28 July 2003; doi:10.1038/nature01973.

- Lewis, E. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
- Struhl, G. Genes controlling segmental specification in *Drosophila* thorax. *Proc. Natl Acad. Sci. USA* **79**, 7380–7384 (1982).
- Hunt, P. & Krumlauf, R. Deciphering the Hox code: clues to patterning branchial regions of the head. *Cell* **66**, 1075–1078 (1991).
- Hammerschmidt, M. *et al.* Mutations affecting morphogenesis during gastrulation and tail formation in the zebrafish *Danio rerio*. *Development* **123**, 143–151 (1996).
- Mlodzik, M., Fjose, A. & Gehring, W. J. Isolation of *caudal*, a *Drosophila* homeo box-containing gene with maternal expression whose transcripts form a concentration gradient at the pre-blastoderm stage. *EMBO J.* **4**, 2961–2969 (1985).
- Katsuyama, Y., Sato, Y., Wada, S. & Saiga, H. Ascidian tail formation requires *caudal* function. *Dev. Biol.* **213**, 257–268 (1999).
- Edgar, L. G., Carr, S., Wang, H. & Wood, W. B. Zygotic expression of the *caudal* homolog *pal-1* is required for posterior patterning in *Caenorhabditis elegans* embryogenesis. *Dev. Biol.* **229**, 71–88 (2001).
- Subramanian, V., Meyer, B. I. & Gruss, P. Disruption of the murine homeobox gene *cdx1* affects axial skeletal identities by altering the mesodermal expression domains of *Hox* genes. *Cell* **83**, 641–653 (1995).
- Chawengsaksophak, K., James, R., Hammond, V. E., Kontgen, F. & Beck, F. Homeosis and intestinal tumours in *cdx2* mutant mice. *Nature* **386**, 84–87 (1997).
- van den Akker, E. *et al.* *Cdx1* and *Cdx2* have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development* **129**, 2181–2193 (2002).
- Beck, F., Chawengsaksophak, K., Waring, P., Playford, R. J. & Furness, J. B. Reprogramming of intestinal differentiation and intercalary regeneration in *Cdx2* mutant mice. *Proc. Natl Acad. Sci. USA* **96**, 7318–7323 (1999).
- Tamai, Y. *et al.* Colonic hamartoma development by anomalous duplication in *Cdx2* knockout mice. *Cancer Res.* **59**, 2965–2970 (1999).
- Charité, J. *et al.* Transducing positional information to the *Hox* genes: critical interaction of *cdx* gene products with position-sensitive regulatory elements. *Development* **125**, 4349–4358 (1998).
- Hunter, C. P., Harris, J. M., Maloof, J. N. & Kenyon, C. *Hox* gene expression in a single *Caenorhabditis elegans* cell is regulated by a *caudal* homolog and intercellular signals that inhibit *Wnt* signaling. *Development* **126**, 805–814 (1999).
- Owens, B. M. & Hawley, R. G. HOX and non-HOX homeobox genes in leukemic hematopoiesis. *Stem Cells* **20**, 364–379 (2002).
- Krumlauf, R. *Hox* genes in vertebrate development. *Cell* **78**, 191–201 (1994).
- Gering, M., Rodaway, A. R. E., Göttgens, B., Patient, R. K. & Green, A. R. The *SCL* gene specifies haemangioblast development from early mesoderm. *EMBO J.* **17**, 4029–4045 (1998).
- Sauvageau, G. *et al.* Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations *in vitro* and *in vivo*. *Genes Dev.* **9**, 1753–1765 (1995).
- Antonchuk, J., Sauvageau, G. & Humphries, R. K. HOXB4-induced expansion of adult hematopoietic stem cells *ex vivo*. *Cell* **109**, 39–45 (2002).
- Buske, C. *et al.* Deregulated expression of HOXB4 enhances the primitive growth activity of human hematopoietic cells. *Blood* **100**, 862–868 (2002).
- Björnsson, J. M. *et al.* Reduced proliferative capacity of hematopoietic stem cells deficient in *hoxb3* and *hoxb4*. *Mol. Cell Biol.* **23**, 3872–3883 (2003).
- Thorsteinsdóttir, U. *et al.* Overexpression of the myeloid leukemia-associated *Hoxa9* gene in bone marrow cells induces stem cell expansion. *Blood* **99**, 121–129 (2002).
- Perkins, A. C. & Cory, S. Conditional immortalization of mouse myelomonocytic, megakaryocytic and mast cell progenitors by the *Hox-2.4* homeobox gene. *EMBO J.* **12**, 3835–3846 (1993).
- Chen, F., Greer, J. & Capecchi, M. R. Analysis of *Hoxa7/Hoxb7* mutants suggests periodicity in the generation of the different sets of vertebrae. *Mech. Dev.* **77**, 49–57 (1998).
- Kappen, C. Disruption of the homeobox gene *Hoxb-6* in mice results in increased numbers of early erythrocyte progenitors. *Am. J. Hematol.* **65**, 111–118 (2000).
- Lawrence, H. J. *et al.* Mice bearing a targeted interruption of the homeobox gene *HOXA9* have defects in myeloid, erythroid, and lymphoid hematopoiesis. *Blood* **89**, 1922–1930 (1997).
- Chase, A. *et al.* Fusion of *ETV6* to the caudal-related homeobox gene *CDX2* in acute myeloid leukemia with the t(12;13)(p13;q12). *Blood* **93**, 1025–1031 (1999).
- Kyba, M., Perlingeiro, R. C. & Daley, G. Q. *HoxB4* confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* **109**, 29–37 (2002).
- Joly, J. S. *et al.* Expression of a zebrafish *caudal* homeobox gene correlates with the establishment of posterior cell lineages at gastrulation. *Differentiation* **50**, 75–87 (1992).
- Kingsley, P. D. *et al.* Subtractive hybridization reveals tissue-specific expression of *ahna* during embryonic development. *Dev. Growth Differ.* **43**, 133–143 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank members of the Zon laboratory, B. Paw and S. Orkin for critical reading of this manuscript. We also thank K. Humphries for *HoxB4* retrovirus, H. G. Frohnhöfer for *kcg* mutants, J. Postlethwait and A. Amores for genomic sequences, and members of the zebrafish community for gifts of cDNAs. L.I.Z. is an Investigator of the Howard Hughes Medical Institute. This work was supported by Legal Sea Foods, the Grousebeck family and grants from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.I.Z. (zon@enders.tch.harvard.edu).

Chemokine receptor CXCR4 downregulated by von Hippel–Lindau tumour suppressor pVHL

Peter Staller^{1*}, Jitka Sulitkova¹, Joanna Lisztwan¹, Holger Moch², Edward J. Oakeley¹ & Wilhelm Krek^{1*}

¹Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland

²Institute of Pathology, University of Basel, Schönbeinstrasse 40, CH-4031 Basel, Switzerland

* Present address: Institute of Cell Biology, ETH Höggerberg, HPM F 42, CH-8093 Zürich, Switzerland

Organ-specific metastasis is governed, in part, by interactions between chemokine receptors on cancer cells and matching chemokines in target organs. For example, malignant breast cancer cells express the chemokine receptor CXCR4 and commonly metastasize to organs that are an abundant source of the CXCR4-specific ligand stromal cell-derived factor-1 α (ref. 1). It is still uncertain how an evolving tumour cell is reprogrammed to express CXCR4, thus implementing the tendency to metastasize to specific organs. Here we show that the von Hippel–Lindau tumour suppressor protein pVHL negatively regulates CXCR4 expression owing to its capacity to target hypoxia-inducible factor (HIF) for degradation under normoxic conditions. This process is suppressed under hypoxic conditions, resulting in HIF-dependent CXCR4 activation. An analysis of clear cell renal carcinoma that manifests mutation of the *VHL* gene in most cases revealed an association of strong CXCR4 expression with poor tumour-specific survival. These results suggest a mechanism for CXCR4 activation during tumour cell evolution and imply that *VHL* inactivation acquired by incipient tumour cells early in tumorigenesis confers not only a selective survival advantage but also the tendency to home to selected organs.

Inactivation of the *VHL* tumour suppressor gene is linked to the development of several different tumour types in humans, including hereditary and sporadic clear cell carcinoma of the kidney². The best characterized function of pVHL is as a recognition subunit of an E3 ubiquitin protein ligase complex that targets the α -subunits of the DNA-binding transcription factor HIF (ref. 3) for ubiquitin-mediated degradation in the presence of oxygen⁴. Tumour-derived pVHL mutants are defective in this regard and manifest constitutive activation of HIF target genes^{5,6}. Indeed, HIF activation is an early event in the evolution of neoplastic kidney lesions in *VHL* disease⁷. To identify novel pVHL-regulated genes, we compared the gene expression profile of pVHL-deficient renal cell carcinoma (RCC) cells previously stably transfected with an empty vector (designated A498(neo)) with those engineered to stably produce haemagglutinin (HA)-tagged wild-type pVHL₃₀ (designated A498(HA-pVHL₃₀)) using Affymetrix gene chip technology. This approach identified 101 genes that were significantly upregulated in pVHL-expressing A498 cells relative to vector control cells, and 64 that were significantly downregulated (Supplementary Fig. S1).

We were particularly intrigued by the fact that, among the genes most strongly suppressed by the reintroduction of functional pVHL₃₀ in A498 cells is the gene encoding the G-protein-coupled chemokine receptor CXCR4 (Supplementary Fig. S1), the receptor for the chemokine stromal cell-derived factor-1 α (SDF-1 α ; also referred to as CXCL12)^{8,9}. Chemokine receptor–ligand interactions have been implicated in the homing of various subsets of haematopoietic cells to specific anatomical sites^{10,11} and to determine the metastatic destination of malignant breast cancer cells¹. The latter finding provided molecular support for the ‘chemoattraction theory’ of organ-specific metastasis^{12,13}. A key unanswered question

concerns the molecular mechanism through which cells acquire CXCR4 during the evolution of tumour cells. The fact that the gene on top of the ‘list’ of pVHL-suppressed genes was CXCR4 provided us with an unexpected avenue down which to pursue this question.

To validate the oligonucleotide microarray results, we performed Northern blot analysis. pVHL-negative A498(neo) cells express high levels of CXCR4 messenger RNA (Fig. 1a, lane 1), whereas their pVHL-expressing counterparts, A498(HA-pVHL₃₀), do not (lane 2). The disappearance of CXCR4 mRNA from pVHL-expressing cells was correlated with a downregulation of HIF2 α protein expression. A498(neo) cells also exhibited strong surface expression of the CXCR4 receptor as demonstrated by immunofluorescence microscopy (Fig. 1b, upper panel). Thus, pVHL₃₀ negatively regulates, directly or indirectly, CXCR4 mRNA and protein production in RCC cells.

The above experiments were performed with an A498 clone expressing pVHL₃₀. To test whether pVHL₁₉ behaves similarly and to assess whether certain naturally occurring tumour-derived mutants of pVHL are defective in CXCR4 regulation, we infected A498 cells with empty control retrovirus or with retroviruses encoding untagged pVHL₃₀, pVHL₁₉ or the HIF degradation-defective tumour mutant pVHL₃₀(N78S)¹⁴. Expression of pVHL₃₀ or pVHL₁₉ suppressed CXCR4 mRNA production (Fig. 1c, lanes 2 and 3), whereas expression of pVHL₃₀(N78S) did not (Fig. 1c, lane 4). pVHL₃₀ mutants that exerted residual degradation activity towards HIF2 α such as pVHL₃₀(Y98H) or pVHL₃₀(R167W)¹⁴ had intermediate suppressive effects on CXCR4 mRNA abundance (data not shown). These results suggest that pVHL is a negative regulator of CXCR4 gene expression in RCC cells and indicate that this function is affected in tumour-derived mutants of pVHL that fail to regulate the HIF system. Moreover, they imply a link between loss of function of pVHL, the ensuing expression of HIF α and the overproduction of CXCR4 mRNA.

In keeping with the above, CXCR4 mRNA was strongly induced in pVHL-positive human embryonic kidney (HEK-293) cells and primary human proximal renal tubular epithelia cells (RPTECs) when they were subjected to low oxygen concentrations (1%) (Fig. 1d, e, respectively). In addition, the hypoxia-mimetic substance cobalt chloride induced CXCR4 mRNA in RPTECs (Fig. 1e). The kinetics of CXCR4 mRNA accumulation in response to hypoxia closely followed that of GLUT3, an established HIF target¹⁵ (Fig. 1e). Taken together, these results suggest that CXCR4 is a hypoxia-inducible gene.

To substantiate this conclusion we cloned the human CXCR4 promoter including the first intronic region. Sequence analysis of this region revealed four potential hypoxia-response elements (HREs) located within 2.6 kilobases upstream of the transcriptional start site and one at position +1 kb within the intron (Fig. 2a). A luciferase reporter containing this CXCR4 promoter fragment, when introduced into *VHL*-positive HEK-293 cells, was activated about twofold by hypoxia (data not shown). Co-transfection of wild-type HIF1 α enhanced hypoxia-inducible reporter activity (about 10-fold; Fig. 2a) and a prolyl-hydroxylation-defective mutant of HIF1 α , HIF1 α (P564A), which escapes pVHL control, activated the CXCR4 promoter also under normoxic conditions (about 7.5-fold; Fig. 2a). A progressive deletion analysis of the CXCR4 promoter revealed that the HRE located at –1.3 kb upstream of the transcriptional start is critical for hypoxia/HIF1 α -inducible reporter activity (Fig. 2a), and mutation of this sequence in the context of the full-length CXCR4 promoter rendered the promoter insensitive to hypoxia and HIF1 α (Fig. 2a). Finally, an oligonucleotide comprising this sequence (referred to hereafter as CXCR4-HRE (CHRE)) bound baculovirus-produced HIF1 α –ARNT (Aryl hydrocarbon receptor nuclear translocator) heterodimers in electrophoretic mobility-shift assays (Fig. 2b, lane 3). Oligonucleotide binding of HIF1 α –ARNT complexes was specific, as shown by competition and supershift experiments

(Fig. 2b, lanes 4–12 and 13–15, respectively). Similarly, whole cell extracts of A498 cells gave rise to a prominent gel-shift complex (Fig. 2c, lane 2), which competed specifically with wild-type but not with a mutant CHRE oligonucleotide (Fig. 2c, lanes 6–8 and 9–11, respectively). In addition, anti-HIF2 α antibodies abolished (Fig. 2c, lane 12) or supershifted (lanes 13 and 14) this complex, suggesting that it contains HIF2 α . Anti-HIF1 α antibodies had no effect (Fig. 2c, lane 15). These results strongly suggest that *CXCR4* is a novel target gene of the DNA-binding transcriptional activator HIF.

Next we asked whether the presence of *CXCR4* on the surface of A498 cells would confer chemoattraction. As shown in Fig. 3a (left panel), stimulation of A498 cells with increasing concentrations of SDF-1 α resulted in directed migration of these cells in a trans-well assay. This response was abrogated by the restoration of wild-type pVHL (Fig. 3a, right panel). Because directional cell migration and tissue invasion require changes in the dynamics of the actin cytoskeleton, we next tested whether *CXCR4* signalling in RCC cells results in the activation of downstream effectors such as LIM kinase 1 (LIMK-1)¹⁶. The latter is important in controlling actin dynamics through inactivation of the actin depolymerization factor cofilin¹⁷. Stimulation of A498 cells with SDF-1 α led to a rapid activation of LIMK-1 (Fig. 3b, left panel) and extracellular signal-related kinases (ERKs), which regulate cell proliferation and the cell's motility machinery (Fig. 3c, upper panel). Restoration of pVHL function abrogated these responses (Fig. 3b, c). The loss of pVHL tumour suppressor function in RCC cells is therefore intimately linked to the acquisition of a novel property, namely the enhanced responsiveness of such cells to chemotactic signals.

Next we asked whether *CXCR4* is overexpressed in human renal tumours and, if so, whether its upregulation is correlated with the inactivation of *VHL*. As a surrogate of the latter, we determined the expression of the HIF target genes encoding carbonic anhydrase (CA9) and glucose transporter 1 (*GLUT1*) because upregulation of these genes is correlated with *VHL* status in clear cell RCC. Expression levels of *CXCR4*, CA9 and *GLUT1* were assessed in 29 clear cell RCC and 5 papillary RCC using real-time quantitative polymerase chain reaction (PCR) and compared with those in normal renal tissue ($n = 7$). As illustrated in Fig. 4a, mRNA levels

Table 1 Cox analysis for clear cell renal carcinoma

Variable	Relative risk	P
<i>CXCR4</i> expression	1.8	<0.01
Differentiation grade	1.9	<0.001
Tumour stage	3.0	<0.0001

of *CXCR4*, CA9 and *GLUT1* were significantly higher in clear cell than in papillary RCC ($P < 0.05$; Supplementary Table S3) or normal renal tissue. Because papillary RCCs are not characterized by *VHL* inactivation, these findings argue that *CXCR4* expression follows *VHL* inactivation and HIF target gene activation in clear cell RCC.

To evaluate the expression of *CXCR4* in a wide range of RCC, we screened a renal cancer tissue microarray (TMA) containing 532 different elements for *CXCR4* protein expression by immunohistochemistry. The antibody used in this study detects high levels of *CXCR4* expression in breast cancer tissue (Fig. 4b), consistent with earlier results. Glomeruli and tubules in normal kidney tissue stained weakly, whereas intensive staining of lymphocellular infiltrates was observed (Fig. 4c). With this antibody we observed a pattern of *CXCR4* expression on the TMA ranging from absent/weak to strong (Fig. 4d–g, respectively). To determine whether *CXCR4* expression is correlated with clinical aggressiveness in clear cell RCC, we focused our analysis on 195 of 407 clear cell carcinoma samples on the TMA for which tumour-specific survival data were available. Within this cohort, strong *CXCR4* expression was detected in 93 samples (47.6%). In the remainder of the TMA elements, staining was either weak or absent. Statistical analysis of these data revealed no significant correlation of high levels of *CXCR4* expression with tumour stage and/or differentiation grade (Supplementary Table S4). However, we found a striking positive correlation between strong *CXCR4* expression and poor tumour-specific survival ($P = 0.001$) in a univariate analysis (Fig. 4h). Importantly, multivariate Cox proportional hazards analysis suggests that this correlation was independent of tumour stage and differentiation grade (Table 1). No correlation with tumour-specific survival was observed for CA9: it displayed strong staining

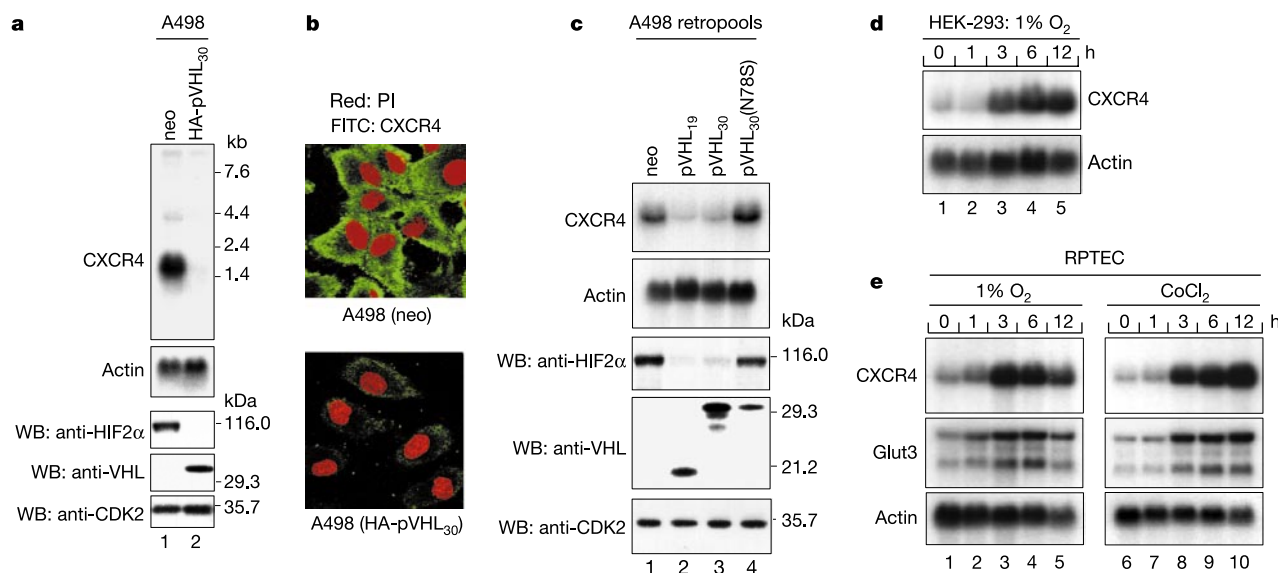


Figure 1 pVHL and hypoxia regulate *CXCR4* expression. **a**, Panels from the top: *CXCR4* mRNA expression in A498(neo) and A498(HA-pVHL₃₀) cells, and immunoblots of HIF2 α , pVHL and CDK2. **b**, A498(neo) and A498(HA-pVHL₃₀) cells stained with anti-*CXCR4* monoclonal antibody (green) and propidium iodide (red). **c**, Retrovirally infected A498 cell

pools stably expressing pVHL₃₀, pVHL₁₉ or pVHL₃₀(N78S) were analysed for the expression of *CXCR4* mRNA and for HIF2 α , pVHL and CDK2 proteins. **d**, **e**, *CXCR4* and *GLUT3* mRNA expression in HEK-293 cells (**d**) and RPTECs (**e**) treated with hypoxia or cobalt chloride (100 μ M) as shown. WB, western blot.

in 169 (87%) of the 195 samples (data not shown). CA9 expression was also not correlated with tumour stage and grade (data not shown). These results imply that a high level of CXCR4 expression is a predictor of poor tumour-specific survival, which in turn suggests that monitoring CXCR4 expression in patients with RCC may provide additional prognostic information.

The data presented in this report suggest a potential mechanism of how and when evolving clear cell RCC cells may acquire CXCR4 and thus the tendency to metastasize to specific secondary sites. It is known that the survival of patients with clear cell RCC is limited by the development of metastasis. The association of strong CXCR4 expression with poor tumour-specific survival implies that CXCR4 expression levels influence the metastatic behaviour of clear cell RCC. The acquisition of CXCR4 expression is likely to occur during the initial phases of tumour progression, set off by the functional inactivation of pVHL. Whether CXCR4 has additional roles, such as the promotion of cell proliferation, that could confer a selective advantage for the tumour cell at the primary site remains to be investigated. Finally, activation of HIF is also the result of intratumoral hypoxia and non-*VHL* oncogenic pathways^{18–20} (for example,

in glioblastoma the expression of CXCR4 is localized to regions of necrosis and angiogenesis²¹). Hence, the mechanism underlying tumour cell-specific production of CXCR4 proposed here may apply to solid tumours in general.

One of the most remarkable features of metastatic tumours is the degree to which they differ genotypically and phenotypically from their primary tumour. A key question that follows is whether these changes have already taken place in the primary tumour, allowing it to spread to a specific secondary site, or whether primary tumour cells that are carried to secondary organs undergo these changes after they have been exposed to the new environment. The results presented here suggest that the expression of a receptor important

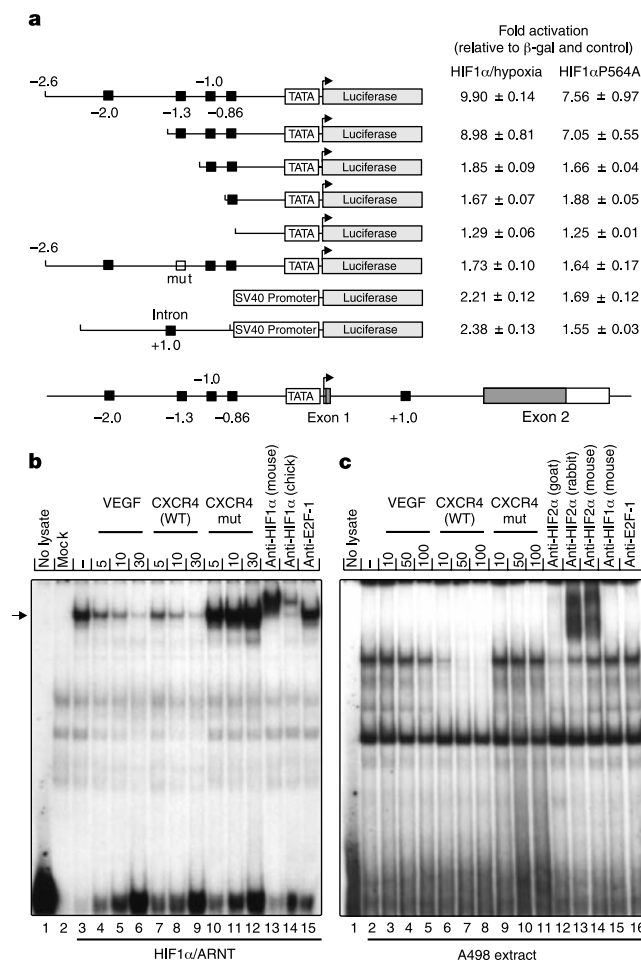


Figure 2 CXCR4 is a target gene of HIF. **a**, Luciferase activity after co-transfection of indicated CXCR4 reporter plasmids and HIF1α alleles in HEK-293 cells. Numbers indicate fold activation relative to β-galactosidase (β-gal) standard and control. Black squares indicate potential HREs. Lower panel: schematic representation of the human CXCR4 locus. **b**, Gel-shift assay using a probe spanning the CHRE and cell extracts from baculovirus-infected Sf9 cells coexpressing human HIF1α and ARNT. Competitor oligonucleotides (amounts in nanograms) and antibodies used are indicated. **c**, Extracts of A498 cells were assayed for CHRE DNA binding. Competition and supershift experiments are indicated as described for **b**. Arrows in **b** and **c** indicate specific DNA binding complexes consisting of HIF1α/ARNT (**b**) and HIF2α/ARNT (**c**). WT, wild type.

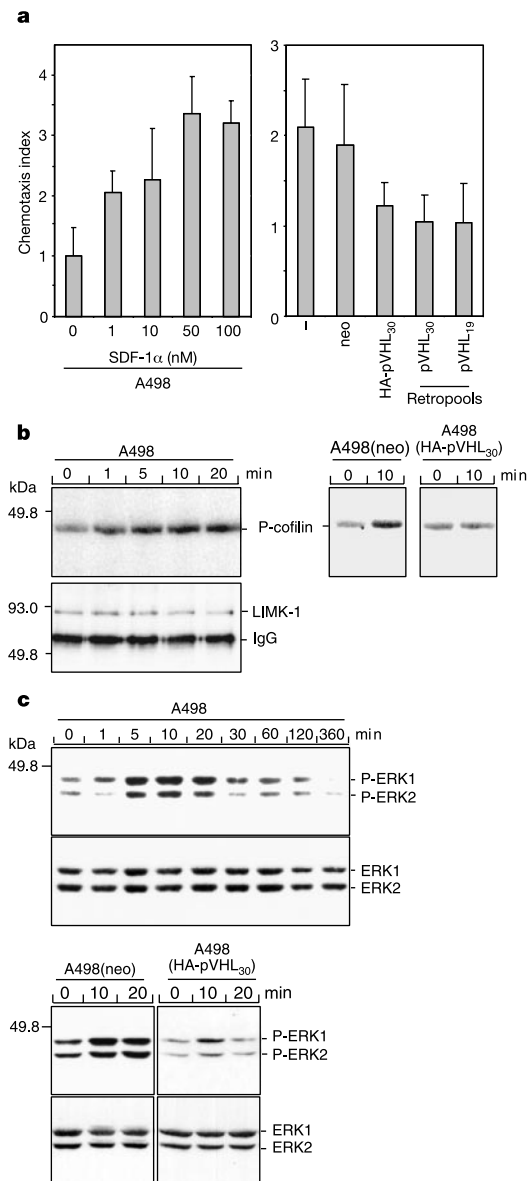


Figure 3 pVHL suppresses SDF-1α-mediated chemotaxis of RCC cells. **a**, Left: chemotactic response of A498 cells to different concentrations of SDF-1α. Results are expressed as the ratio between cells migrating towards the chemokine gradient and cells migrating in the negative control. Right: chemotaxis of A498 cells expressing the indicated pVHL species after exposure to 10 nM SDF-1α. **b**, *In vitro* LIM kinase-1 assay after treatment of A498, A498(neo) and A498(HA-pVHL₃₀) cells with SDF-1α. **c**, Cells were treated for indicated durations with 10 nM SDF-1α and processed for immunoblotting with anti-phospho-ERK1/ERK2 specific antibodies (upper panels) or anti-ERK1/2 total protein antibodies (lower panels).

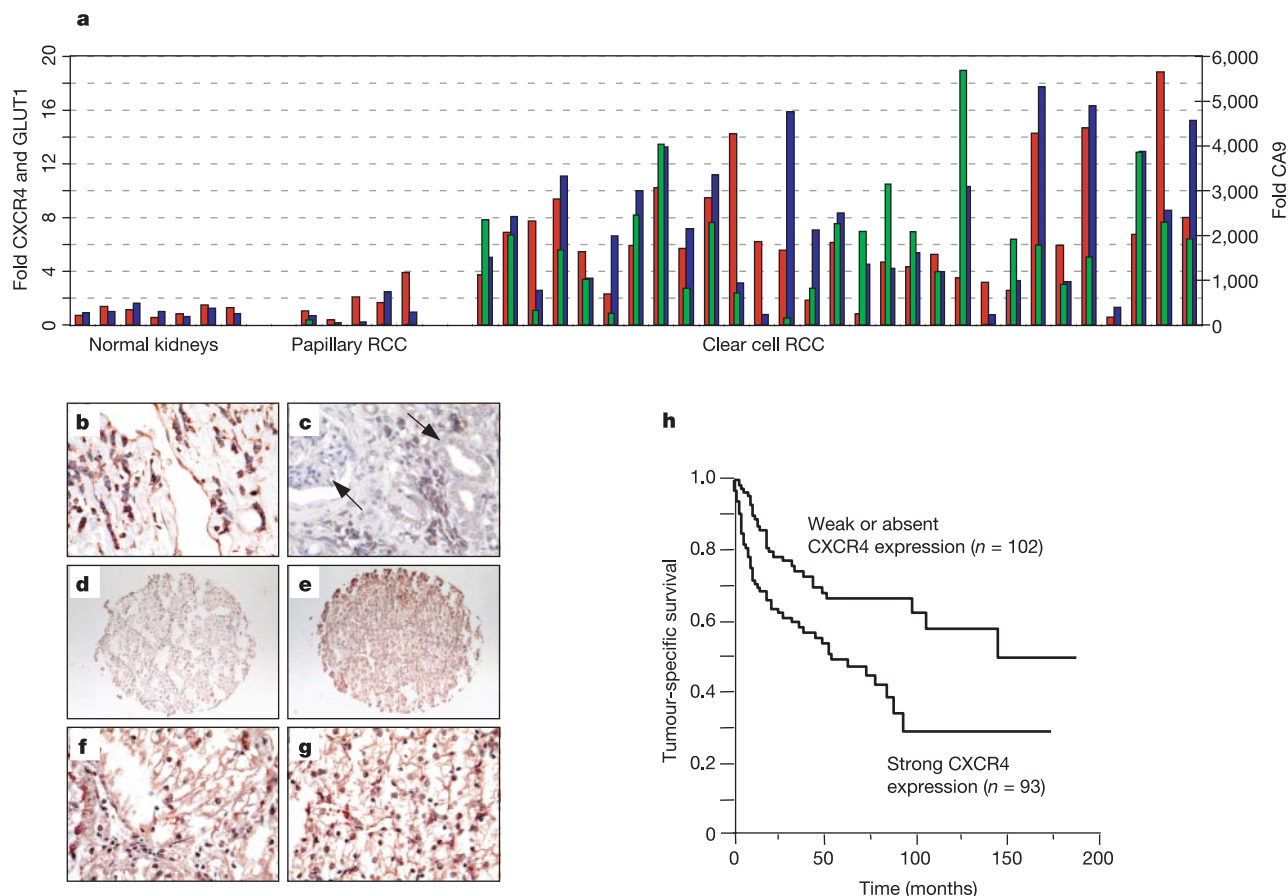


Figure 4 CXCR4 expression in human renal cell carcinoma. **a**, Quantitative reverse transcriptase PCR analysis of normal kidney tissues and renal cell carcinoma. Fold expression levels are normalized to 18S rRNA and compared with the average values in normal kidneys. Red bars, CXCR4; blue bars, GLUT1; green bars, CA9. **b-g**, Immunohistochemical evaluation of CXCR4 expression. **b**, Invasive breast cancer.

for homing to distant organs is directly regulated by the hypoxia pathway and that its upregulation is already seen in early tumour stages. The propensity to metastasize, at least for certain cancers, may therefore be determined by the identities of the mutant alleles acquired relatively early during multistep tumorigenesis²². □

Methods

Cell culture, retroviral infections and reporter-gene assays

Cell culture conditions and procedures for retroviral infections as well as plasmid constructions are available from the authors on request. A 2.7-kb fragment of the human CXCR4 promoter and the intronic region was amplified by PCR from A498 genomic DNA and subcloned into pGL2b and pGL2-promoter vector (Promega), respectively. HEK-293 cells were subjected to 12 h of hypoxia in a three-gas incubator (Nabco) 24 h after transfection. Luciferase activity was corrected for β -galactosidase activity and at least three independent experiments were performed.

Gene microarray analysis

Analysis was performed with HG_U95A GeneChips (Affymetrix). Scanning was done in an Affymetrix GeneChip scanner. Analysis was performed using Affymetrix Microarray Suite v5 and GeneSpring 4.2.1 (Silicon Genetics). Changes in gene expression were assessed by looking for concordant changes between replicates by using a signed Wilcoxon rank test. The 'change' threshold P was less than 0.003 for increase and more than 0.997 for decrease. After concordance analysis these values become less than 9×10^{-6} and more than 0.999991 respectively. Any gene whose detection P was more than 0.05 in all experimental conditions was discarded from the analysis. Cluster analysis was performed with dCHIP software (www.biostat.harvard.edu/complab/dchip).

Immunoblotting and northern analysis

Immunoblot analysis for HIF2 α (EPAS1-ab199; Novus Biologicals), pVHL²³, CDK2 (M2; Santa Cruz), ERK (Cell Signaling Technology catalogue no. 9102) and phospho-ERK (Cell

Signaling Technology catalogue no. 9101) were performed with standard techniques. Total RNA was prepared with TRIzol reagent (Life Technologies). Blots were hybridized in ExpressHyb (Clontech) to ³²P-labelled probes.

DNA binding assays

Preparation of cell lysates and binding reactions were performed as described^{24,25}. Oligonucleotide sequences for vascular endothelial growth factor and CXCR4 are available from the authors on request. Antibodies used were: H1 α 67 (Novus Biologicals), affinity-purified anti-human HIF1 α chicken polyclonal antibody (gift from M. Gassmann), anti-HIF2 α (EPAS1-ab199, HIF2 α -ab8365, Novus Biologicals; C-16, Santa Cruz) and anti-E2F-1 (KH-95, Santa Cruz).

Chemotaxis assays

Cell migration was assayed with trans-well inserts with 8- μ m pore membranes. Human SDF-1 α was obtained from R&D Systems. Membranes were precoated with fibronectin (0.5 μ g ml⁻¹). After starvation for 24 h, cells were detached with PBS containing 2.5 mM EDTA, centrifuged and resuspended in DMEM, 0.1% BSA, 12 mM HEPES pH 7.4 at 2×10^5 cells per 100 μ l. After 4 h, cells on the lower surface were fixed and stained with crystal violet. Five different fields of the membrane were counted to assess the number of cells that had migrated.

In vitro kinase reactions

Polyclonal rabbit serum to human LIMK-1 (anti-LIMK-1) was raised against a fusion protein of full-length LIMK-1 and maltose-binding protein (J.L., unpublished work) and affinity-purified²³. A498 cells were serum-starved for 24 h, treated with 1 nM SDF-1 α for the indicated times, subjected to immunoprecipitation as described²⁶ and processed for kinase reactions in the presence of glutathione S-transferase-cofilin¹⁶.

Quantitative PCR

Total RNA was extracted from frozen tissue, treated with DNase I using the RNeasy system (Qiagen) and reverse transcribed with random hexamer primers (Amersham) and AMV reverse transcriptase (Roche). Complementary DNA was analysed by the fluorogenic

5'-nuclease PCR assay. Primers and probes (Applied Biosystems): CXCR4 forward 5'-TGGGTGGTTGGTTCCAGTTT-3', reverse 5'-ATGCAATAGCAGGACAGGATGA-3', probe 5'-FAM-CATGGTTGGCCTTATCTGCCTGGTA-TAMRA-3' (where TAMRA represents carboxytetramethylrhodamine). GLUT1, CA9 and 18S rRNA were detected with Assays-on-Demand primer and probe sets Hs00197884_m1, Hs00154208_m1 and Hs99999901_s1, respectively (Applied Biosystems). Gene-specific PCR products were measured continuously by an ABI PRISM 7000 Sequence Detection System (Applied Biosystems) during 40 cycles. 18S rRNA was used for normalization. Expression values for tumour samples were compared with the expression in normal renal tissue. Student's *t*-test was applied to analyse the expression between different grades, tumour stages and tumour types using StatView 5.0 PPC (SAS Inst. Inc., Cary, North Carolina). *P* < 0.05 was considered as significant.

Immunofluorescence and tissue microarray analysis

Indirect immunofluorescence staining of CXCR4 was performed as described²⁷ with MAB173 (R&D Systems). Confocal laser scanning microscopy was performed on a FV500/BX61 microscope (Olympus). Immunohistochemistry was performed on a tissue microarray containing 532 renal tumours and normal tissue²⁸ obtained from the renal nephrectomy series at the University of Basel²⁹. Approval to analyse the tissues was obtained from the ethical commission in Basel. Upon antigen retrieval, standard avidin-biotin complex immunohistochemistry (ABC-Elite; Vectra Laboratories) was used to evaluate CXCR4 (MAB 172; R&D Systems; dilution 1:2000) and CA9 (M75 (ref. 30), dilution 1:160) expression. Colour was developed with 3-amino-9-ethylcarbazole and sections were counterstained with haematoxylin. CXCR4 and CA9 positivity was assessed semiquantitatively by staining intensity as negative, weak or strongly positive by two independent investigators without prior knowledge of the clinical follow-up data. Cases were accepted only as strongly positive if reviewers independently defined them thus. Contingency table analysis was used to analyse the relationship of CXCR4 and CA9 expression to differentiation grade and tumour stage. Clear cell RCCs were analysed for patient survival with the Kaplan-Meier method. Statistical differences between the groups were determined with the log-rank test. A Cox proportional hazards analysis was used to test for independent prognostic information. Tumour-specific survival data were obtained by reviewing the hospital records, by direct communication with the attending physicians, and from the Cancer Registry of Basel. Patients were evaluated from the time of biopsy diagnosis to the last known follow-up. Tumour-specific clinical follow-up data were available from 195 patients.

Received 8 May; accepted 26 June 2003; doi:10.1038/nature01874.

- Müller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
- Kaelin, W. G. Jr Molecular basis of the VHL hereditary cancer syndrome. *Nature Rev. Cancer* **2**, 673–682 (2002).
- Semenza, G. L. HIF-1 and tumor progression: Pathophysiology and therapeutics. *Trends Mol. Med.* **8**, S62–S67 (2002).
- Pugh, C. W. & Ratcliffe, P. J. The von Hippel-Lindau tumor suppressor, hypoxia-inducible factor-1 (HIF-1) degradation, and cancer pathogenesis. *Semin. Cancer Biol.* **13**, 83–89 (2003).
- Iliopoulos, O., Levy, A. P., Jiang, C., Kaelin, W. G. Jr & Goldberg, M. A. Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. *Proc. Natl Acad. Sci. USA* **93**, 10595–10599 (1996).
- Wiesener, M. S. *et al.* Constitutive activation of hypoxia-inducible genes related to overexpression of hypoxia-inducible factor-1 α in clear cell renal carcinomas. *Cancer Res.* **61**, 5215–5222 (2001).
- Mandriota, S. J. *et al.* HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. *Cancer Cell* **1**, 459–468 (2002).
- Bleul, C. C. *et al.* The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829–833 (1996).
- Oberlin, E. *et al.* The CXCR4 chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* **382**, 833–835 (1996).
- Forster, R. *et al.* CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. *Cell* **99**, 23–33 (1999).
- Peled, A. *et al.* Dependence of human stem cell engraftment and repopulation of NOD/SCID mice on CXCR4. *Science* **283**, 845–848 (1999).
- Chambers, A. E., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nature Rev. Cancer* **2**, 563–572 (2002).
- Liotta, L. A. An attractive force in metastasis. *Nature* **410**, 24–25 (2001).
- Clifford, S. C. *et al.* Contrasting effects on HIF-1 α regulation by disease-causing pVHL mutations correlate with patterns of tumorigenesis in von Hippel-Lindau disease. *Hum. Mol. Genet.* **10**, 1029–1038 (2001).
- Maxwell, P. H. *et al.* Hypoxia-inducible factor-1 modulates gene expression in solid tumors and influences both angiogenesis and tumor growth. *Proc. Natl Acad. Sci. USA* **94**, 8104–8109 (1997).
- Nishita, M., Aizawa, H. & Mizuno, K. Stromal cell-derived factor 1 α activates LIM kinase 1 and induces cofilin phosphorylation for T-cell chemotaxis. *Mol. Cell. Biol.* **22**, 774–783 (2002).
- Arber, S. *et al.* Regulation of actin dynamics through phosphorylation of cofilin by LIM-kinase. *Nature* **393**, 805–809 (1998).
- Aebbersold, D. M. *et al.* Expression of hypoxia-inducible factor-1 α : A novel predictive and prognostic parameter in the radiotherapy of oropharyngeal cancer. *Cancer Res.* **61**, 2911–2916 (2001).
- Harris, A. L. Hypoxia—a key regulatory factor in tumour growth. *Nature Rev. Cancer* **2**, 38–47 (2002).
- Semenza, G. L. Signal transduction to hypoxia-inducible factor 1. *Biochem. Pharmacol.* **64**, 993–998 (2002).
- Rempel, S. A., Dudas, S., Ge, S. & Gutierrez, J. A. Identification and localization of the cytokine SDF1 and its receptor, CXCR4 chemokine receptor 4, to regions of necrosis and angiogenesis in human glioblastoma. *Clin. Cancer Res.* **6**, 102–111 (2000).
- Bernards, R. & Weinberg, R. A. A progression puzzle. *Nature* **418**, 823 (2002).
- Hergovich, A., Lisztwan, J., Barry, R., Ballschmieter, P. & Krek, W. Regulation of microtubule stability by the von Hippel-Lindau tumour suppressor protein pVHL. *Nature Cell Biol.* **5**, 64–70 (2003).

- Takahashi, Y., Takahashi, S., Shiga, Y., Yoshimi, T. & Miura, T. Hypoxic induction of prolyl 4-hydroxylase α (I) in cultured cells. *J. Biol. Chem.* **275**, 14139–14146 (2000).
- Wang, G. L. & Semenza, G. L. Purification and characterization of hypoxia-inducible factor 1. *J. Biol. Chem.* **270**, 1230–1237 (1995).
- Lisztwan, J. *et al.* Association of human CUL-1 and ubiquitin-conjugating enzyme CDC34 with the F-box protein p45(SKP2): Evidence for evolutionary conservation in the subunit composition of the CDC34-SCF pathway. *EMBO J.* **17**, 368–383 (1998).
- Amara, A. *et al.* HIV coreceptor downregulation as antiviral principle: SDF-1 α -dependent internalization of the chemokine receptor CXCR4 contributes to inhibition of HIV replication. *J. Exp. Med.* **186**, 139–146 (1997).
- Moch, H. *et al.* High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am. J. Pathol.* **154**, 981–986 (1999).
- Moch, H. *et al.* Prognostic utility of the recently recommended histologic classification and revised TNM staging system of renal cell carcinoma: A Swiss experience with 588 tumors. *Cancer* **89**, 604–614 (2000).
- Pastorekova, S., Zavadova, Z., Kostal, M., Babusikova, O. & Zavada, J. A novel quasi-viral agent, MaTu, is a two-component system. *Virology* **187**, 620–626 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank all members of the laboratory for discussions; K. Struckmann for providing the tumour RNA samples; G. Keller and B. Mohr for human *GLUT3* and *CXCR4* cDNAs, respectively; and R. Bernards, U. Muller, N. Hynes, G. Thomas and members of our laboratory for critically reading the manuscript. This work was supported by the Robert Wenner Award, the Dr Josef Steiner Foundation and the Novartis Research Foundation. H.M. is supported by a grant from the Swiss National Science Foundation. P.S. is supported by an EMBO long-term fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.K. (wilhelm.krek@cell.biol.ethz.ch). The data for the microarray experiment are deposited in ArrayExpress under accession number E-MEXP-13.

The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase

Lionel Pintard^{1*}, John H. Willis^{2*}, Andrew Willems³,
Jacque-Lynne F. Johnson⁴, Martin Srayko^{4†}, Thimo Kurz², Sarah Glaser¹,
Paul E. Mains⁴, Mike Tyers³, Bruce Bowerman² & Matthias Peter¹

¹Institute of Biochemistry ETH, Hönggerberg 8093, Zürich, Switzerland

²Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403, USA

³Department of Medical Genetics and Microbiology, Samuel Lunenfeld Research Institute, Mt Sinai Hosp., 600 University Ave, Univ. Toronto, Toronto, Ontario, M5G 1X5, Canada

⁴Genes and Development Research Group and Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, T2N 4N1, Canada

* These authors contributed equally to this work

† Present address: Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

Many biological processes, such as development and cell cycle progression are tightly controlled by selective ubiquitin-dependent degradation of key substrates. In this pathway, the E3-ligase recognizes the substrate and targets it for degradation by the 26S proteasome. The SCF (Skp1-Cul1-F-box) and ECS (Elongin C-Cul2-SOCS box) complexes are two well-defined cullin-based E3-ligases^{1–3}. The cullin subunits serve a scaffolding function and interact through their C terminus with the RING-finger-containing protein Hrt1/Roc1/Rbx1, and through their N terminus with Skp1 or Elongin C, respectively. In *Caenorhabditis elegans*, the ubiquitin-ligase activity of the CUL-3 complex is required for degradation of the microtubule-severing protein MEI-1/katanin at the meiosis-to-mitosis transition⁴. However, the molecular composition of this cullin-based E3-ligase is not known. Here

we identified the BTB-containing protein MEL-26 as a component required for degradation of MEI-1 *in vivo*. Importantly, MEL-26 specifically interacts with CUL-3 and MEI-1 *in vivo* and *in vitro*, and displays properties of a substrate-specific adaptor. Our results suggest that BTB-containing proteins may generally function as substrate-specific adaptors in Cul3-based E3-ubiquitin ligases.

Structural analysis of SCF and ECS complexes revealed that Skp1 and Elongin C share a common BTB-domain fold, which interacts with the N-terminal parts of cullins Cul1 and Cul2^{1–3}. BTB domains are named for the *Drosophila melanogaster* transcription factors, Bric-a-brac (bab), Tramtrack (ttk), and Broad-Complex (BR-C)⁵. These observations raise the possibility that BTB-containing proteins may associate with cullins, and therefore define a previously unknown class of E3-ligases. In *C. elegans*, the ubiquitin-ligase activity of a CUL-3 complex is required to degrade MEI-1 after meiosis, which is essential for the assembly of a functional mitotic spindle^{4,6,7}. MEI-1 is the subunit of the katanin-like microtubule-severing heterodimer MEI-1/MEI-2 that localizes to the spindle and chromosomes during meiosis⁸. Neddylation and deneddylation both positively regulate the activity of the CUL-3 complex *in vivo*⁴, but the molecular composition of this CUL-3-based E3-ligase is not known.

To identify components of the *C. elegans* CUL-3 ubiquitin-ligase, we took a genetic approach and screened for conditional *mel* (maternal-effect lethal) mutants that displayed similar phenotypes to those caused by reducing the function of CUL-3 or the NED-8 conjugation pathway⁴. We identified a recessive, temperature-sensitive, maternal-effect embryonic lethal mutation termed *or543ts*. Time-lapse recording of *or543ts* mutant embryos during mitosis using Nomarski differential interference (DIC) optics revealed severe defects in spindle positioning, spindle elongation,

and cytokinesis (Fig. 1). Microtubules examined by indirect immunofluorescence appeared shorter and disorganized (Supplementary Fig. 1).

All together, these defects were strikingly reminiscent of the phenotypes observed in *cul-3(RNAi)* embryos, and in a mutant of the ubiquitin-activating enzyme of the NED-8 conjugation pathway, *rfl-1(or198ts)* (ref. 6). *rfl-1(or198ts)* and *cul-3(RNAi)* embryos are defective for MEI-1 degradation: therefore, we asked whether reducing *mei-1* function by RNAi in *or543ts* mutant embryos could rescue the observed defects in mitotic spindle positioning and cytokinesis. Indeed, *or543ts; mei-1(RNAi)* double mutant embryos were able to correctly position the mitotic spindle and terminate cytokinesis (Fig. 2a, b). Thus, the *or543* mutation prevents down-regulation of MEI-1/2 katanin after meiosis.

Genetic mapping of *or543ts* placed it in the region of a previously characterized gene called *mel-26*^{9,10}. MEL-26 has been described as a postmeiotic inhibitor of the MEI-1/2 complex^{7,9,10}, suggesting that MEL-26 may be part of a pathway controlling MEI-1 degradation. To address this question, we monitored MEI-1 expression in *mel-26(RNAi)* embryos using a MEI-1::GFP line. In contrast to the wild type, MEI-1-GFP accumulated at the forming mitotic spindle in *mel-26(RNAi)* embryos (Fig. 2c, arrowheads). Furthermore western blot experiments revealed an increase in MEI-1-GFP levels in *mel-26(RNAi)* embryos, suggesting that MEI-1 regulation by MEL-26 occurs at the level of protein stability (Fig. 2d). Finally, endogenous MEI-1 strongly accumulated to similar levels in *cul-3* and *mel-26(RNAi)* embryos (Fig. 2e). We conclude that both MEL-26 and CUL-3 control the degradation of MEI-1 after meiosis.

Although neddylation and deneddylation of CUL-3 is essential for the ligase activity of the CUL-3 complex⁴, western blot analysis revealed that CUL-3 neddylation was unaffected in two *mel-26* mutants (*or543* and *or184*) or in *mel-26(RNAi)* embryos (Fig. 2f).

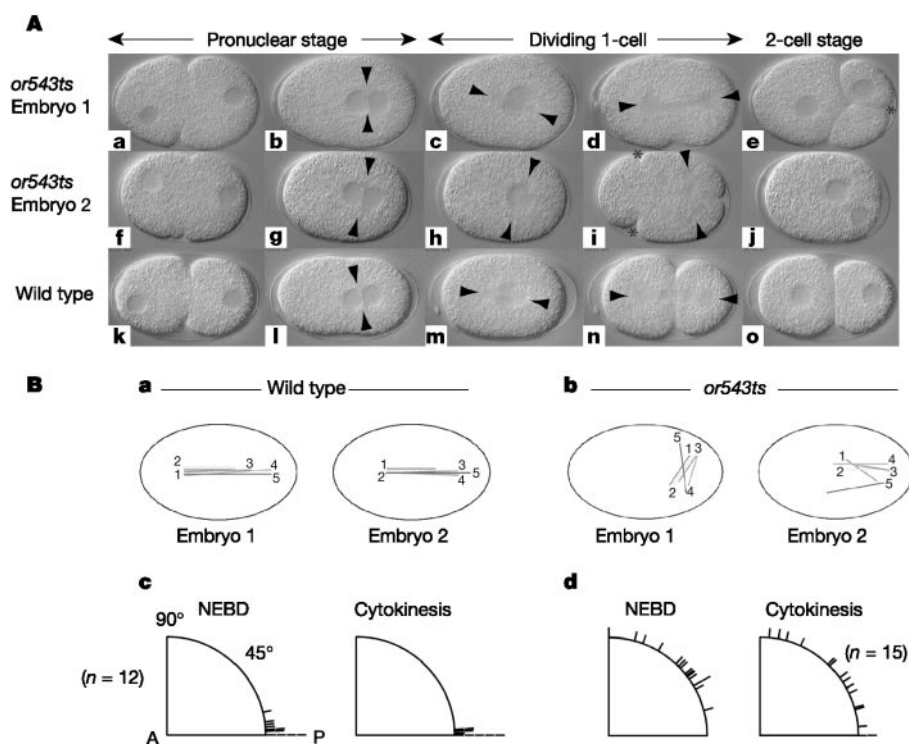


Figure 1 *or543ts* embryos display defects in microtubule-dependent processes. **A**, The first two cell divisions in two (*or543ts*) mutant embryos compared to the wild type. **A**, **a–e**, Mutant embryo 1. **A**, **f–j**, Mutant embryo 2. **A**, **k–o**, Wild-type embryos. Arrowheads mark the position of the centrosomes; asterisks the position of ectopic furrows. **B**, Schematic representation of the mitotic spindle position at nuclear envelope breakdown (NEBD), and

during cytokinesis (**B**, **c**, **d**) in *or543ts* and wild-type embryos. The position of the mitotic spindle in individual embryos is depicted at 1 min intervals. **B**, **a**, Wild-type embryo 1 and 2; **B**, **b**, *or543ts* embryo 1 and 2. Alignment along the anterior–posterior axis corresponds to 0° degrees. Each black bar in **B**, **c**, **d** corresponds to one embryo.

Likewise, in contrast to *rfl-1(or198ts)* mutants, NED-8 localized to the nucleus and the cytoplasm in wild-type and *mel-26(or543)* mutant embryos (Fig. 2g). These observations imply that MEL-26 acts downstream or in parallel to the NED-8 conjugation pathway.

MEL-26 contains a BTB domain, raising the possibility that it might directly interact with CUL-3. To determine whether MEL-26 and CUL-3 form a complex *in vivo*, we prepared *C. elegans* embryo extracts and separated the proteins on sucrose gradients (5–20%). Western blot analysis with anti-CUL-3 and anti-MEL-26 antibodies revealed that a fraction of MEL-26 cosedimented with CUL-3 at approximately 7.3S (corresponding to 150 kDa) (data not shown). To determine whether CUL-3 and MEL-26 indeed interact *in vivo*, we performed co-immunoprecipitation experiments from embryo extracts. Importantly, MEL-26 co-precipitated with CUL-3 (Fig. 3a), and conversely, CUL-3 co-precipitated with MEL-26 (Fig. 3b). Moreover, *in vitro* translated CUL-3 was specifically retained on GST–MEL-26 affinity columns (Fig. 3c). Taken together, these experiments demonstrate that CUL-3 binds MEL-26 *in vivo* and *in vitro*.

In classical SCF complexes, Skp1 binds the N-terminal part of the cullin, suggesting that MEL-26 might similarly bind the N-terminal part of CUL-3. Although the N termini of cullins are poorly conserved, multiple alignments identified three subfamilies (CUL-1, CUL-2/5; CUL-3/4) and revealed several amino acids that are common among them (Fig. 4d). These residues are excellent candidates for mediating specific interaction. In support of this prediction we found that MEL-26 interacted with CUL-3 but not CUL-2 by two-hybrid assay (Fig. 3e). Moreover, deletion of the N-terminal part of CUL-3 abolished the interaction (Fig. 3e), consistent with the prediction that MEL-26 may interact with the

N-terminal part of CUL-3. On the basis of the structure of the SCF^{Skp-2} complex², these conserved residues in Cul3 orthologues cluster on the same surface region as the Skp1 binding site of Cul1. Indeed, mutating two of these highly conserved residues of CUL-3 into alanine (S51AF52A) abolished its interaction with MEL-26, as determined by two-hybrid assay (Fig. 3e). Similarly, mutating the corresponding residues of Cdc53 (Y48 and M49, highlighted in red) abolished the interaction with Skp1 (Supplementary Information; and A.W. and M.T., unpublished results).

We next asked whether the negative regulation of MEI-1/2 by MEL-26 involves a physical interaction. Two-hybrid assays revealed that MEL-26 was able to bind MEI-1 but not its heterodimeric partner MEI-2 (Fig. 4a and Supplementary Fig. 3). Moreover, MEL-26 failed to interact with the protein encoded by a gain-of-function allele, *mei-1(ct46 gf)* (Fig. 4a)⁹. This observation is significant because the *mei-1(ct46 gf)* allele results in ectopic MEI-1 expression during the early cleavages, leading to phenotypes similar to those caused by reducing MEL-26 function⁷. The residue mutated in *mei-1(ct46 gf)* (P99 to L) lies within a predicted high score PEST motif (ref. 11 and Supplementary Information). PEST motifs contribute to the rapid degradation of several short-lived proteins such as cyclins¹². To confirm these observations we performed *in vitro* binding experiments. As shown in Fig. 4b, HA-tagged MEI-1 but not MEI-1(gf) expressed in yeast was able to interact with GST–MEL-26 *in vitro* (Fig. 4b). Thus, these results suggest that recognition of *mei-1(ct46 gf)* by MEL-26 is impaired *in vivo* and support the idea that MEL-26 may be the substrate-specific adaptor of the CUL-3 complex, perhaps analogous to a traditional F-box protein.

Consistent with this possibility, MEL-26 shares other properties

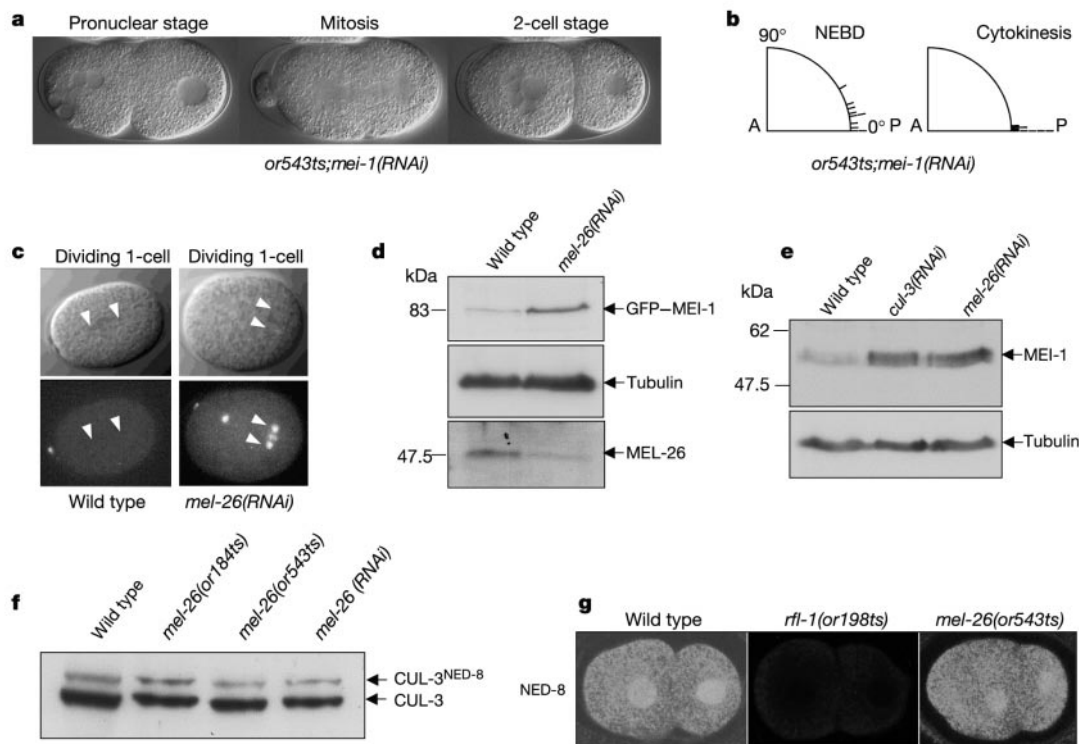


Figure 2 Inactivation of *mei-1* by RNAi suppresses the defects in microtubule-dependent processes of *or543ts* mutant embryos. **a**, DIC images of the *mei-1(RNAi); or543ts* double mutant embryo during pronuclear migration, mitosis and at the two-cell stage. **b**, Schematic representation of the position of the mitotic spindle in the double *or543ts; mei-1(RNAi)* mutant, at nuclear envelope breakdown, and during cytokinesis (scored as described in Fig. 1B). **c**, DIC (upper panel) and GFP images (lower panel) of wild-type (left panel) and *mei-26(RNAi)* embryos (right panel) expressing MEI-1::GFP visualized during

mitosis. White arrows point to the position of centrosomes. **d**, MEI-1::GFP levels in extracts prepared from wild-type and *mei-26(RNAi)* embryos. **e**, MEI-1 levels in extracts prepared from wild-type, *cul-3(RNAi)* and *mei-26(RNAi)* embryos. **f**, Examination of neddylated (upper band) and non-neddylated (lower band) forms of CUL-3 in wild-type, *mei-26* mutant and *mei-26(RNAi)* embryos. **g**, Confocal micrographs of two-cell stage embryos from wild-type, *rfl-1(or198)* and *mei-26(or543ts)* mutant embryos stained with anti-NED-8 antibodies.

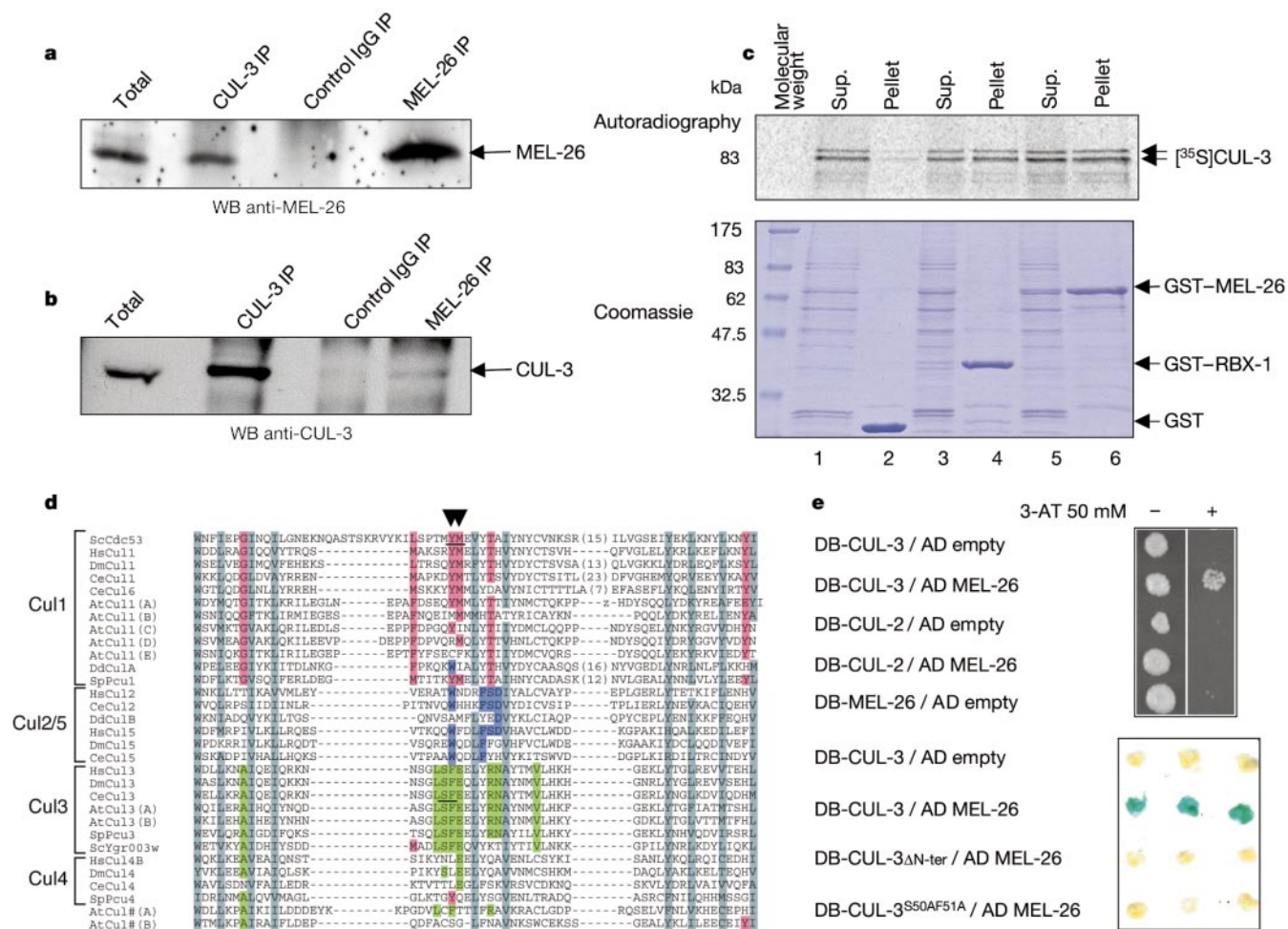


Figure 3 MEL-26 interacts with CUL-3 *in vivo* and *in vitro*. **a, b**, MEL-26 and CUL-3 were immunoprecipitated as indicated with affinity-purified anti-MEL-26, anti-CUL3 or for control rabbit-immunoglobulin IgG antibodies, and the washed immunoprecipitates (IP) were immunoblotted for the presence of MEL-26 (**a**) or CUL-3 (**b**). Note that small amounts of CUL-3 non-specifically bound to control IgG in **b**. **c**, CUL-3 was radiolabelled by *in vitro* translation and incubated with either glutathione-S-transferase (GST), the known cullin binding protein GST-RBX-1 (positive control) or GST-MEL-26 fusion proteins. Bound proteins were eluted and separated on SDS-PAGE ('Pellet' fraction). 1/10th of the supernatant was loaded ('Sup.' fraction) for control. **d**, Multiple alignments of the cullin N termini from different species identified three subfamilies (Cul1, Cul2/5, and Cul3/4). Residues specifically conserved in the Cul1 subfamily are highlighted in red; those conserved in the Cul2/5 and Cul3/4 subfamilies are highlighted in blue and in green

with F-box proteins. It has been reported that F-box proteins are unstable and subject to SCF-mediated ubiquitination by an autocatalytic mechanism^{13–15}. Interestingly, MEL-26 strongly accumulated in *cul-3(RNAi)* embryos (Fig. 4c), suggesting that MEL-26 may be regulated in a similar manner. Furthermore, it has been proposed that F-box proteins dimerize within active SCF complexes^{16,17}. Similarly, we found that MEL-26 interacted with itself and with the product of the dominant-negative allele *mel-26(sb45)*, which was unable to bind MEI-1 (Supplementary Fig. 3). These observations are consistent with the dominant-negative properties of *sb45* (ref. 10) and strongly support the idea of a requirement for MEL-26 homodimerization *in vivo*.

Through the isolation and characterization of a maternal-effect lethal mutant, we identified MEL-26 as a component of the machinery involved in degradation of the putative microtubule-severing protein MEI-1. Three lines of evidence suggest that MEL-26 is a core component of the CUL-3 complex and may

respectively, selected hydrophobic residues conserved in all cullins are coloured in grey for orientation of the alignment. Residues in Cdc53 that are involved in Skp1 binding (Y48 and M49) and residues in CUL-3 (S50 and F51) required for binding to MEL-26 are underlined. **e**, MEL-26 specifically interacts with CUL-3 but not with CUL-2 by two-hybrid assay. Yeast cells carrying the histidine reporter (Y190), transformed with the indicated combination of plasmids were spotted on plates containing '–' (control) or '+' 50 mM 3-AT. Point mutations in the conserved N-terminal part of CUL-3 (S50AF51A) abolished binding to MEL-26. The interaction was determined by two-hybrid analysis on three colonies using the β -galactosidase filter assay. Western blot experiments with anti-CUL-3 antibodies confirmed that wild-type and the mutated version of CUL-3 were expressed at similar levels and at the expected size (Supplementary Fig. S2).

function as a substrate-specific adaptor of the complex. First, genetic and biochemical data suggest that MEL-26 and CUL-3 function as negative regulators of MEI-1. Second, MEL-26 and CUL-3 form a complex *in vivo*. Third, MEL-26 displays the properties of a substrate-specific adaptor, because it specifically interacts with MEI-1 and strongly accumulates after inactivation of *cul-3*. Moreover, like MEL-26, the cullin-binding proteins Skp1 and Elongin C contain a BTB-domain and their interface with the F-box and the SOCS-box adaptors reveal striking structural similarities. Indeed, both interfaces are interchangeable, implying that a topology similar to both interfaces could be achieved by a single polypeptide such as MEL-26, which contains a BTB domain and a MATH protein–protein interaction domain (Supplementary Fig. 3).

Taken together, our data suggest that a single polypeptide, such as MEL-26, that incorporates features of both Skp1/Elongin C proteins (cullin-binding) and the F-box/SOCS box proteins (substrate recognition) may assemble into a previously unknown type of

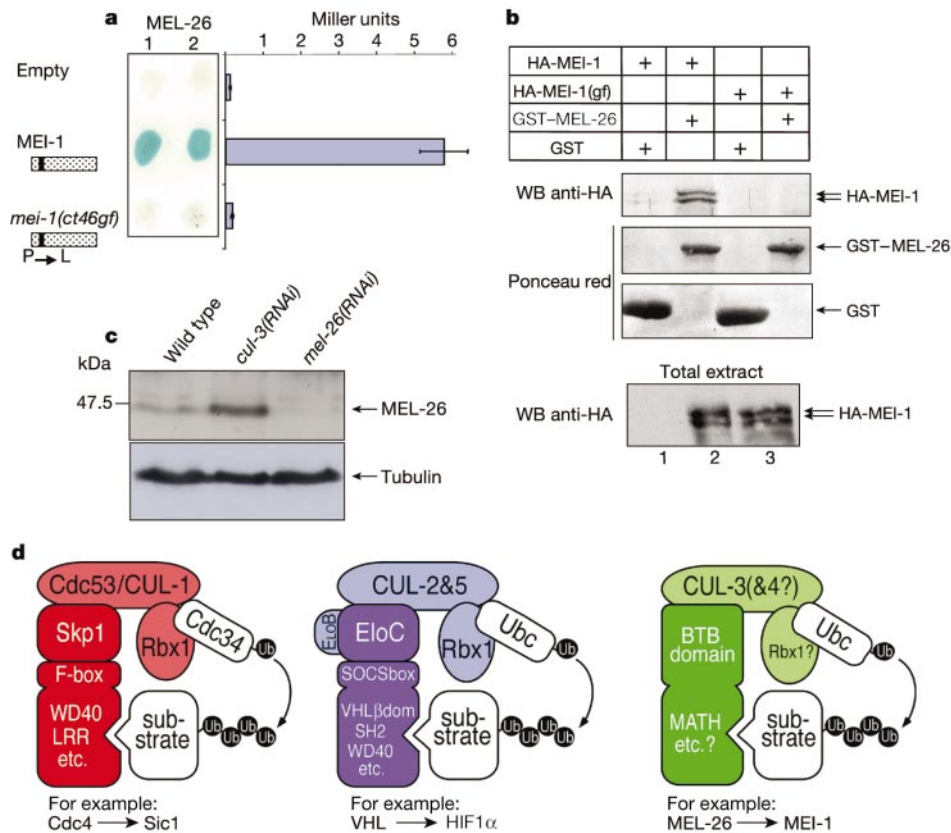


Figure 4 MEL-26 displays the properties of an F-box protein. **a**, Two-hybrid interactions of MEL-26 and wild-type MEI-1 or *mei-1(ct46gf)*. Interactions of the indicated constructs were determined by β -galactosidase assays of two yeast colonies (1 and 2) on filters (left panel) or quantitated by β -galactosidase liquid assays (right panel). The PEST motif of MEI-1 is shown as a black box, the mutated residue in the *mei-1(ct46gf)* allele is indicated below. **b**, GST-MEL-26 binds HA-MEI-1 but not HA-MEI-1(gf) *in vitro*. Yeast extracts containing either HA-MEI-1 or HA-MEI-1(gf) were incubated with GST or GST-MEL-26. The complexes were purified on glutathione-Sepharose 4B and the bound proteins separated on SDS-PAGE. The membrane was stained with ponceau red to reveal GST

proteins (middle panels) and blotted with anti-HA antibodies (upper panel). For control, the bottom panel shows the levels of HA-tagged MEI-1 (lane 2) and MEI-1(gf) (lane 3) in the total extract. An extract without HA-MEI-1 is included in lane 1. **c**, MEL-26 accumulates after inactivation of *cul-3*. The protein level of MEL-26 was analysed by western blot and compared in wild-type, *cul-3(RNAi)* and *mei-26(RNAi)* embryos. **d**, Model showing how a single polypeptide such as MEL-26 may assemble with CUL-3 and define a previously unknown type of E3-ligase as compared to the well-characterized SCF and ECS complexes.

E3-ligase. Furthermore, recent findings from W. Harper's laboratory suggest that BTB-containing proteins may generally function as substrate-specific adaptors of a Cul3-based ubiquitin ligase²⁷. Many proteins containing both a BTB domain and an additional protein-protein interaction domain are involved in the organization of actin and microtubule dynamics, such as kelch¹⁸, mayven¹⁹ or gigaxonin (GAN)²⁰. Gigaxonin is mutated in axonal neuropathy, leading to a disorganization of intermediate filaments. Several identified mutations affect its BTB domain, implying that the BTB domain is functionally relevant. Interestingly, like MEL-26, a yeast BTB protein (YLR108c) is able to specifically interact by two-hybrid assay with the yeast Cul3 homologue encoded by the product of the gene YGR003W (B. Luke and M.P., unpublished results). Thus, we suggest that BTB-containing proteins may generally function as substrate-specific adaptors of a new type of E3-ligase involving Cul3. □

Methods

Strains and alleles

C. elegans Bristol strain N2 was used as the wild type. Nematode culturing and genetics were performed as described²¹. The mutations used are *dpy-5(e61)* I, *mei-1(ct46gf)* I, *mei-26(or543ts)*, *or184ts*, *ct61sb4* and *sb45* I *dpy-11(e224)* V, *dpy-17(e164)* III, *lon-2(e678)* X, *rfl-1(or198ts)* III, *rol-6(e187)* II, *sma-3(e491)* III, *unc-5(e53)* IV, *unc-32(e189)* III, *unc-50(e306)* III. Temperature-sensitive mutant strains were grown at 15°C and shifted to 25°C between 4 and 24 h before the embryos were examined.

Genetic mapping and positional cloning

Linkage group mapping and meiotic mapping were performed using standard methods as described²¹. Linkage group mapping and meiotic recombination mapping placed *or543ts* to approximately +3.46 on LG I. *mei-26* maps to +3.65 on LG I and has similar mitotic spindle defects (16). A complementation test with *mei-26(ct61sb4)* revealed that *or543ts* is an allele of *mei-26*. Subsequent sequence analysis revealed a missense mutation at amino acid number 127 (R to C). In addition, we isolated a second allele called *or184ts*. In this allele, a missense mutation at amino acid number 82 (G to E) was detected. Both mutations are within the predicted MATH domain of MEL-26 (Wormbase).

Molecular biology techniques

Plasmids are listed in Supplementary Table 1. Details of plasmid constructions are available upon request. Standard procedures were used for recombinant DNA manipulations²². MEI-1, *mei-1(ct46gf)* and MEL-26 were amplified by PCR and inserted in the two-hybrid vectors containing the Gal4-activation domain (pACT-2) (Clontech)—GAL4 DNA-AD, *LEU2*, *amp^r*, HA epitope tag; 8.1 kb—as well as in the plasmid containing the Gal4 DNA binding domain (pAS2-1)—GAL4 DB, *TRP1*, *amp^r*, *CYH^r*; 8.4 kb.

Protein extracts, antibodies and immunoblotting

Embryo extracts were prepared by hypochlorite treatment. Adult worms were bleached, washed several times in M9 buffer, resuspended in 1 volume of laemmli buffer 3X and boiled for 5 min at 95°C. Standard procedures were used for SDS-PAGE and western blotting. The following antibodies were used in this study: primary antibodies were anti-MEL-26 (this study), anti-MEI-1 (this study), anti-GFP (Roche), anti-Tubulin (Sigma), and anti-CUL-3 (ref. 4). The vector pGEX-4T (Pharmacia) was used to express GST-fusions of MEL-26 and MEI-1 in *E. coli* BL21. Antibodies were affinity-purified as described. Secondary antibodies conjugated to peroxidase against rabbit or mouse antibodies were purchased from Biorad.

Native protein extracts and co-immunoprecipitations

N2 worms were grown in liquid culture, collected by sedimentation and bleached by hypochlorite treatment. Embryos were collected, washed several times in M9 buffer and snap-frozen in liquid nitrogen. The embryos were resuspended in 5 volumes of extraction buffer (buffer E: 20 mM Hepes, 150 mM NaCl, 2 mM MgCl₂, 10 mM EDTA, 1 mM DTT containing protease inhibitors ("Complete" mix from Roche)) and broken with a one-shot cell disruptor (Constant Systems) set to the maximal pressure (2.8 kbar). Total lysate was collected and clarified by centrifugation (three times for 10 min at 13,000 r.p.m.). The extract was then incubated for 2 h at 4 °C with affinity-purified antibodies cross-linked to protein G-Sepharose (Pharmacia) using dimethylpimelimidate (Sigma). The beads were washed five times with buffer E. Bound proteins were eluted in 75 µl of gel sample buffer. Western blotting was performed with polyclonal anti-CUL-3 and monoclonal anti-MEL-26 antibodies (this study).

In vitro binding assays

In vitro translated CUL-3 was produced with the TNT-reticulocyte lysate system (Promega) using [³⁵S] methionine and [³⁵S] cysteine (Promix from Amersham). GST fusion proteins were produced in the bacteria strain BL21pRep4 and incubated with *in vitro* translated CUL-3 in 200 µl buffer B (binding buffer: 50 mM TRIS/Cl pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 0.2% NP40). After 45 min at room temperature, 20 µl of glutathione-Sepharose 4B was added to the reaction and incubated for an additional 45 min. After several washes with buffer B, bound proteins were resuspended in Laemmli buffer and separated by SDS-PAGE. Proteins were visualized by Coomassie staining, and the dried gel was exposed for autoradiography and bands quantified.

Yeast two-hybrid experiments

The yeast strain Y190 (Clontech) was used for two hybrid experiments²³. Expression of the *HIS3*-reporter was determined by growth on plates containing 0 or 50 mM of 3-amino-triazole (3-AT, Sigma). β-galactosidase assays were performed on nitrocellulose filters or in liquid cultures as described previously²⁴.

Microscopy, immunocytochemistry and GFP imaging

Microscopy experiments were carried out as described previously⁶ on a Zeiss 200M microscope, and photographed with an Orca camera.

Sequence analysis

Searches for potential PEST motifs in MEI-1 were accomplished using the PESTfind website: <http://www.at.embnat.org/embnat/tools/bio/PESTfind/>. A single high score (+8.80) PEST motif was found in MEI-1 between amino-acids 88 and 104 (RQSGSPPEPADPDVWSK). The mutated residue (P99) in *mei-1(ct46 gf)* is underlined, and this decreases the PEST score to +3.24. Database searches were performed using iterated PSI-BLAST searches of GenBank²⁵. Sequence alignments were performed using ClustalX followed by manual adjustment²⁶.

Received 20 June; accepted 26 July 2003; doi:10.1038/nature01959.

Published online 3 September 2003.

- Schulman, B. A. *et al.* Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* **408**, 381–386 (2000).
- Zheng, N. *et al.* Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703–709 (2002).
- Stebbins, C. E., Kaelin, W. G. Jr & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
- Pintard, L. *et al.* Neddylation and deneddylation of CUL-3 is required to target MEI-1/Katanin for degradation at the meiosis-to-mitosis transition in *C. elegans*. *Curr. Biol.* **13**, 911–921 (2003).
- Zollman, S., Godt, D., Prive, G. G., Couderc, J. L. & Laski, F. A. The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*. *Proc. Natl Acad. Sci. USA* **91**, 10717–10721 (1994).
- Kurz, T. *et al.* Cytoskeletal regulation by the Nedd8 ubiquitin-like protein modification pathway. *Science* **295**, 1294–1298 (2002).
- Clark-Maguire, S. & Mains, P. E. Localization of the *mei-1* gene product of *Caenorhabditis elegans*, a meiotic-specific spindle component. *J. Cell Biol.* **126**, 199–209 (1994).
- Srayko, M., Buster, D. W., Bazirgan, O. A., McNally, F. J. & Mains, P. E. MEI-1/MEI-2 katanin-like microtubule severing activity is required for *Caenorhabditis elegans* meiosis. *Genes Dev.* **14**, 1072–1084 (2000).
- Mains, P. E., Kempthues, K. J., Sprunger, S. A., Sulston, I. A. & Wood, W. B. Mutations affecting the meiotic and mitotic divisions of the early *Caenorhabditis elegans* embryo. *Genetics* **126**, 593–605 (1990).
- Dow, M. R. & Mains, P. E. Genetic and molecular characterization of the *Caenorhabditis elegans* gene, *mel-26*, a postmeiotic negative regulator of *mei-1*, a meiotic-specific spindle component. *Genetics* **150**, 119–128 (1998).
- Clark-Maguire, S. & Mains, P. E. *mei-1*, a gene required for meiotic spindle formation in *Caenorhabditis elegans*, is a member of a family of ATPases. *Genetics* **136**, 533–546 (1994).
- Salama, S. R., Hendricks, K. B. & Thorner, J. G1 cyclin degradation: the PEST motif of yeast Cln2 is necessary, but not sufficient, for rapid protein turnover. *Mol. Cell. Biol.* **14**, 7953–7966 (1994).
- Galan, J. M. & Peter, M. Ubiquitin-dependent degradation of multiple F-box proteins by an autocatalytic mechanism. *Proc. Natl Acad. Sci. USA* **96**, 9124–9129 (1999).
- Wirbelauer, C. *et al.* The F-box protein Skp2 is a ubiquitylation target of a Cul1-based core ubiquitin ligase complex: evidence for a role of Cul1 in the suppression of Skp2 expression in quiescent fibroblasts. *EMBO J.* **19**, 5362–5375 (2000).
- Zhou, P. & Howley, P. M. Ubiquitination and degradation of the substrate recognition subunits of SCF ubiquitin-protein ligases. *Mol. Cell* **2**, 571–580 (1998).
- Wolf, D. A., McKeon, F. & Jackson, P. K. F-box/WD-repeat proteins pop1p and Sud1p/Pop2p form complexes that bind and direct the proteolysis of cdc18p. *Curr. Biol.* **9**, 373–376 (1999).

- Seibert, V. *et al.* Combinatorial diversity of fission yeast SCF ubiquitin ligases by homo- and heterooligomeric assemblies of the F-box proteins Pop1p and Pop2p. *BMC Biochem.* **3**, 1–15 (2002).
- Robinson, D. N. & Cooley, L. *Drosophila* kelch is an oligomeric ring canal actin organizer. *J. Cell Biol.* **138**, 799–810 (1997).
- Soltysik-Espanola, M. *et al.* Characterization of Mayven, a novel actin-binding protein predominantly expressed in brain. *Mol. Biol. Cell* **10**, 2361–2375 (1999).
- Bomont, P. *et al.* The gene encoding gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is mutated in giant axonal neuropathy. *Nature Genet.* **26**, 370–374 (2000).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Ausubel, F. M. *et al.* *Current Protocols in Molecular Biology* (Greene Publishing Associates and Wiley-Interscience, New York, 1991).
- Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
- Breeden, L. K. N. Regulation of the yeast HO gene. *Cold Spring Harb. Symp. Quant. Biol.* **50**, 643–650 (1985).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Xu, L. *et al.* BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3. *Nature* advance online publication; 3 September 2003 (doi:10.1038/nature01985).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to the *C. elegans* Genetics Centre (funded by the NIH National Centre for Research Resources) for providing strains. We thank R. Fischer for generation of the monoclonal anti-MEL-26 antibody, P. Gönczy for introducing L.P. to *C. elegans* and for sharing material and reagents, P. Weissert for help with worm liquid cultures, J. M. Bellanger and I. Sumara for suggestions, P. Wiget for help with microscopy, and P. Gönczy for critical reading of the manuscript. L.P. was supported by a Long-Term Fellowship from the Federation of European Biochemical Societies (FEBS) and a Fellowship from Roche, T.K. by a predoctoral fellowship from the American Heart Association, J.H.W. by a NIH Molecular Biology Training Grant, P.E.M. by grants from the Canadian Institutes of Health Research and the Alberta Heritage Foundation for Medical Research, B.B. by the NIH and M.P. by the ETHZ and the Swiss National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.P. (lionel.pintard@bc.biol.ethz.ch) or M.P. (matthias.peter@bc.biol.ethz.ch).

BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3

Lai Xu^{1*}, Yue Wei^{2*}, Jerome Reboul^{5†}, Philippe Vaglio⁵, Tae-Ho Shin⁴, Marc Vidal⁵, Stephen J. Elledge^{1,3,†} & J. Wade Harper^{1,2,†}

¹Verna and Marrs McLean Department of Biochemistry and Molecular Biology, ²Department of Molecular Physiology and Biophysics, ³Howard Hughes Medical Institute, ⁴Department of Cell and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

⁵Dana Farber Cancer Institute, Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

* These authors contributed equally to this work

† Present addresses: Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA (J.W.H.); Center for Genetics and Genomics, Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA (S.J.E.); and INSERM Unite 119, Institut Paoli Calmette, 13009 Marseille, France (J.R.)

Programmed destruction of regulatory proteins through the ubiquitin-proteasome system is a widely used mechanism for controlling signalling pathways^{1,2}. Cullins³ are proteins that function as scaffolds for modular ubiquitin ligases typified by the SCF (Skp1-Cul1-F-box) complex^{4–6}. The substrate selectivity of these E3 ligases is dictated by a specificity module that binds cullins. In the SCF complex, this module is composed of Skp1, which binds directly to Cul1, and a member of the F-box family of

proteins⁴⁻⁷. F-box proteins bind Skp1 through the F-box motif⁷, and substrates by means of carboxy-terminal protein interaction domains^{1,2,5}. Similarly, Cul2 and Cul5 interact with BC-box-containing specificity factors through the Skp1-like protein elongin C². Cul3 is required for embryonic development in mammals and *Caenorhabditis elegans*⁸⁻¹⁰ but its specificity module is unknown. Here we report the identification of a large family of BTB-domain proteins as substrate-specific adaptors for *C. elegans* CUL-3. Biochemical studies using the BTB protein MEL-26 and its genetic target MEI-1 (refs 12, 13) indicate that BTB proteins merge the functional properties of Skp1 and F-box proteins into a single polypeptide.

Multicellular eukaryotes contain at least six distinct cullin genes³, which are implicated in protein ubiquitination^{1,2}. The core

E3 ubiquitin ligase module is composed of a highly conserved C-terminal cullin homology domain, which is covalently modified by the activator Nedd8, and is associated with the essential Ring-H2 protein Rbx1, which binds E2 ubiquitin-conjugating enzymes (reviewed in refs 1, 2). Substrate choice is dictated by the identity of the specificity module tethered to the cullin amino terminus. F-box and BC-box proteins, the receptors for Cul1 and Cul2/5-based E3s, have been implicated in the destruction of a large number of regulatory proteins^{1,2}. The Cul3 ubiquitin ligase is poorly understood by comparison. Mammalian Cul3 has been implicated in turnover of monomeric forms of cyclin E, and mice lacking *Cul3* die at about embryonic day 6.5 (refs 8, 9). In *C. elegans*, loss of *cul-3* leads to defects in spindle positioning in single-cell embryos, resulting in failed cytokinesis¹⁰. Recent studies^{10,11} suggest that

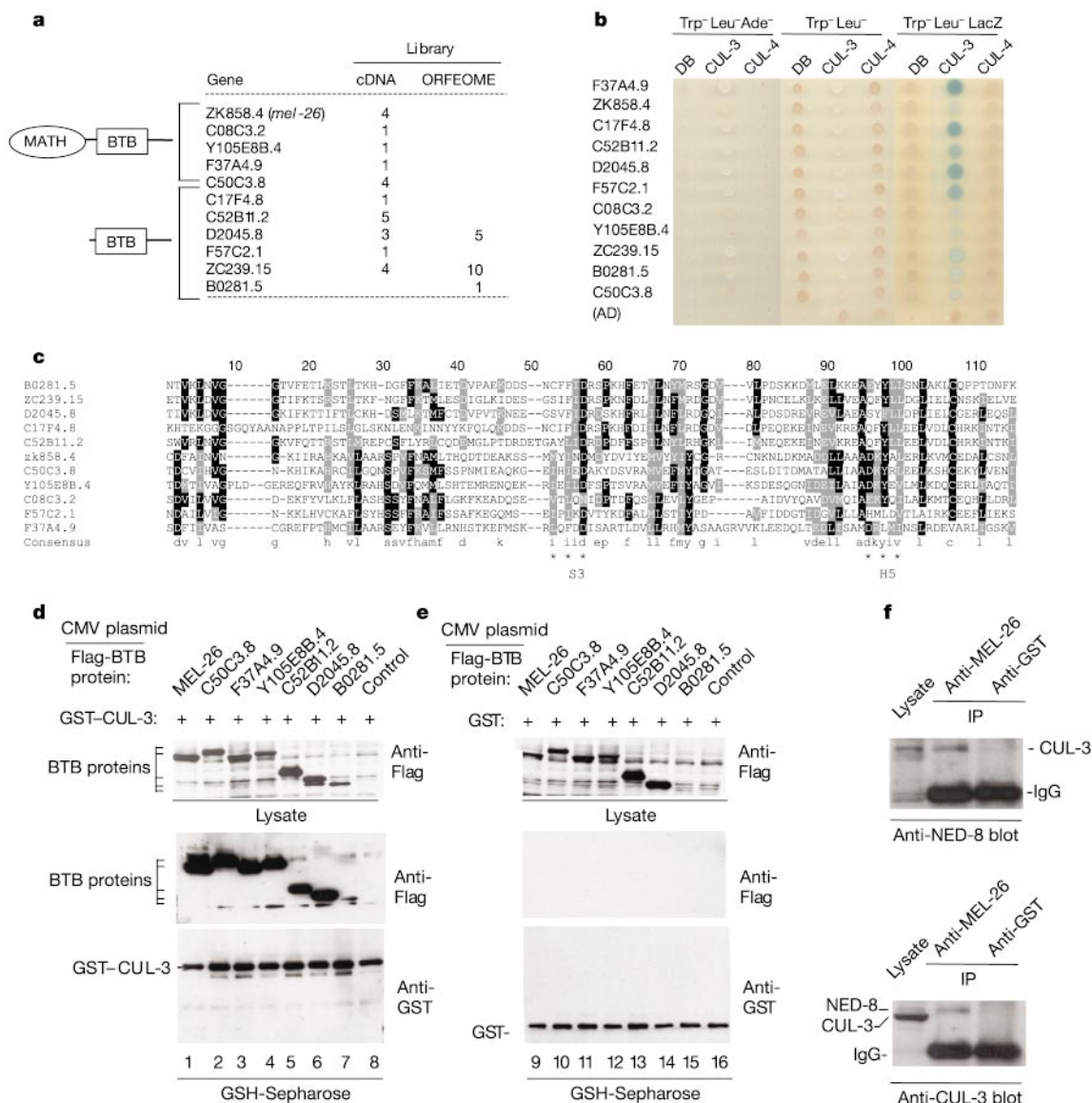


Figure 1 Interaction of *C. elegans* CUL-3 with BTB-domain-containing proteins. **a**, Summary of CUL-3-interacting proteins identified using the yeast two-hybrid system. The number of clones isolated is shown. **b**, Specificity of CUL-3-BTB protein interactions. Two-hybrid assays were performed on cells expressing the indicated proteins. Growth in the absence of adenine (Ade⁻) denotes an interaction, as do blue colonies. **c**, Clustal W alignment of BTB domains from CUL-3-interacting proteins. The positions of critical residues in S3 and H5 of MEL-26 are indicated by an asterisk. **d**, **e**, Interaction of BTB

proteins with CUL-3 in tissue culture cells. Thirty-six hours after transfection with the indicated plasmids, GST proteins were purified with GSH-Sepharose and complexes immunoblotted with anti-Flag and anti-GST antibodies. **f**, Association of CUL-3 and MEL-26 in *C. elegans* embryo extracts. Anti-MEL-26 immune complexes were immunoblotted with anti-CUL-3 (lower panel), the blot stripped and re-probed with anti-NED-8 antibodies (upper panel).

CUL-3 functions to regulate the abundance of MEI-1, a protein required for meiotic spindle function but which is inhibitory to mitotic spindle function^{12,13}.

The mechanism used by Cul3 to recognize substrates such as MEI-1 is unknown, but given the structural conservation with Cul1 (ref. 4), we suspected an SCF-like specificity module. To search for such a module, we screened *C. elegans* two-hybrid complementary DNA libraries and the ORFOME library¹⁴ for CUL-3-interacting proteins. We identified 11 different gene products that could mediate a two-hybrid interaction with CUL-3 but not with CUL-4 (Fig. 1a, b). All of them encoded proteins that contained a conserved BTB/POZ domain (Fig. 1c). This domain, which was initially identified in the *Drosophila* transcriptional repressors *broad complex*, *tramtrack* and *bric-a-brac*, is present in more than 120 *C. elegans* and 190 human proteins (reviewed in ref. 15; J.W.H.,

unpublished data). Interestingly, Skp1 and the BTB domain of human promyelocytic leukaemia zinc finger (PLZF) protein adopt similar three-dimensional α/β -structures^{16,17}, suggesting structural conservation of the interaction with cullins. A subset of BTB proteins that we identified in the screen additionally contain a meprin and TRAF homology (MATH) domain¹⁸.

To verify an interaction between CUL-3 and BTB proteins, we used a transient expression system in 293T cells. All seven BTB proteins tested were found to interact efficiently with glutathione S-transferase (GST)–CUL-3 but not with GST alone (Fig. 1d, e). These data indicate that CUL-3 is capable of interacting with a variety of BTB proteins, analogous to Cul1, which interacts with various F-box proteins. Of the BTB proteins identified, only one (ZK858.4/*mel-26*) has been characterized previously. Mutations in *mel-26* result in embryonic lethality and defective mitotic spindle

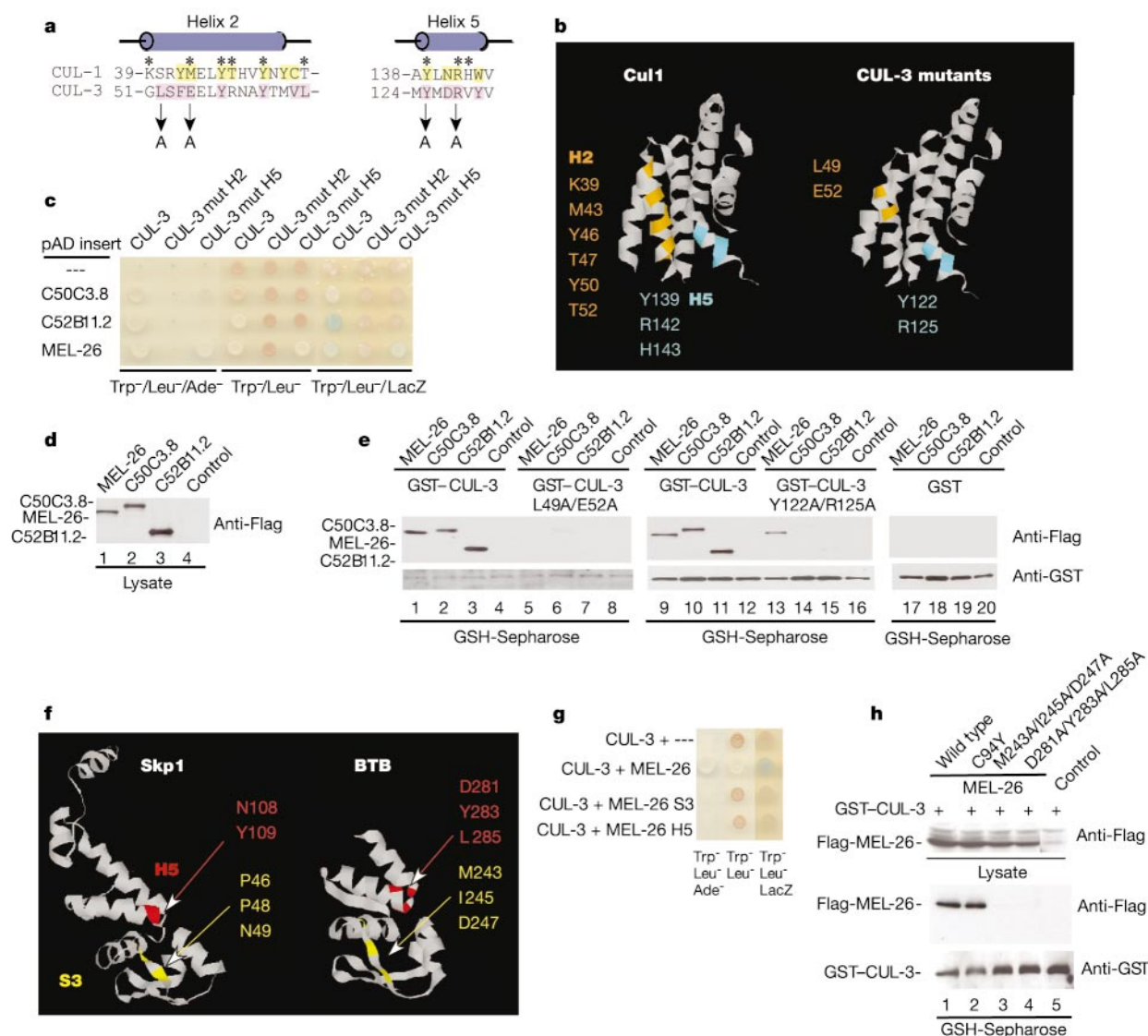


Figure 2 Sequence requirements for interaction of BTB-domain-containing proteins with CUL-3. **a**, Conserved residues in helix 2 (H2) and helix 5 (H5) of Cul1 (yellow) and Cul-3 (pink). **b**, Positions of residues in H2 and H5 of Cul1 that interact with Skp1, and the position of corresponding residues in CUL-3 used to generate H2 (L49/E52) and H5 (Y122/R125) mutants. **c–e**, Mutations in H2 and H5 of CUL-3 block interaction with BTB-domain proteins C50C3.8, C52B11.2 and MEL-26 in the two-hybrid system (**c**) and in

transfected 293T cells (**d, e**). Transfected proteins were assayed either directly in extracts (**d**) or after purification on GSH-Sepharose (**e**). **f**, Structural relationship between S3 and H5 in Skp1 and the analogous surface of the BTB domain of PLZF. **g, h**, The indicated compound point mutations in the predicted S3 and H5 elements of MEL-26 block binding to CUL-3 in the two-hybrid system (**g**) and in transiently transfected tissue culture cells (**h**).

function during the first mitotic division¹², a phenotype similar to that seen in *C. elegans* Nedd8 (*ned-8*) pathway mutants and with *cul-3(RNAi)*¹⁰. To verify physical interaction between MEL-26 and CUL-3 *in vivo*, embryonic extracts were subjected to immunoprecipitation with anti-MEL-26 antibodies and blots were probed sequentially with anti-CUL-3 and anti-NED-8 antibodies. Both CUL-3 and NED-8 were found in MEL-26 immune complexes but not in control complexes (Fig. 1f). As expected, approximately 5% of CUL-3 exists in the more slowly migrating NED-8-modified form¹¹, yet MEL-26 immune complexes contain primarily NED-8-modified CUL-3. Thus, MEL-26 interacts with an activated form of CUL-3 in *C. elegans* embryo extracts, consistent with genetic requirements for *ned-8*. The accompanying paper also found an association between MEL-26 and CUL-3 in embryo extracts¹⁹.

We next investigated the structural relationship between Skp1–Cul1 and BTB–Cul3 interfaces. Two α -helical elements in the N terminus of Cul1—helix 2 (H2) and helix 5 (H5)—contain residues that contact Skp1, and several residues in these helices, including K39, M43, Y139 and R142, are invariant in Cul1 homologues (Fig. 2a, b; see also ref. 4). Structural analysis of CUL-3 indicates that it also contains highly conserved N-terminal helical elements corresponding to H2 and H5 in Cul1 proteins (Fig. 2a, b; see also ref. 4). To examine the involvement of these elements in BTB binding, we generated CUL-3 mutants in H2 (L49A/E52A) and H5 (Y122A/R125A), and tested them for interaction with two MATH-BTB proteins (MEL-26 and C50C3.8) and the BTB-only protein C52B11.2. Mutation of H2 in CUL-3 abolished interaction with all three BTB proteins in both the two-hybrid system and in transfected 293T cells (Fig. 2c–e). By contrast, mutations in H5 have more complex phenotypes. Although interaction with C52B11.2 and C50C3.8 was markedly affected by mutation of H5, the interaction with MEL-26 was largely unaffected (Fig. 2c–e).

Both Skp1 and elongin C adopt α/β -structures similar to the BTB domain from human PLZF (Fig. 2f; see also ref. 16). Several residues in Skp1 located in β -strand 3 (S3) and in α -helix 5 (H5) contact Cul1 (ref. 4). We generated mutations in analogous structural elements of MEL-26 (Figs 1c and 2f) and tested these for interaction with CUL-3. In yeast and in 293T cells, mutation of either S3 or H5 abolished association with CUL-3 (Fig. 2g, h). Taken together, these data indicate that analogous structural elements in Skp1 and BTB proteins are used to interact with a common interface located in the N terminus of distinct cullin proteins. The extent of sequence identity among BTB proteins (about 30–40%) mirrors that seen with the F-box motif⁷.

The fact that multiple BTB-domain proteins interact with CUL-3

and that many BTB-domain proteins also contain additional protein interaction domains points to the possibility that BTB proteins function directly as adaptors that recruit substrates to CUL-3, thereby forming an SCF-like E3. To examine this question, we performed a two-hybrid screen with the MATH-domain-containing MEL-26 protein (Fig. 3a). This screen identified several candidate interacting proteins, including MEL-26 itself, as well as MEI-1, a protein previously demonstrated to be negatively regulated by MEL-26 (ref. 12). MEI-1, an AAA-type ATPase, functions with MEI-2 to form *C. elegans* katanin, a complex that possesses microtubule-severing activity required for meiosis²⁰. As the embryo exits meiosis and enters the first mitotic division, MEI-1 is eliminated through a mechanism that requires *cul-3*, *ned-8* and the COP9/signalosome, suggesting a ubiquitin-dependent process^{9,10}. This process is required to establish mitotic cycles. Although *mel-26* is also required for MEI-1 elimination¹², its biochemical role in this process is unknown. Two heterologous systems were used to examine interaction between MEL-26 and MEI-1. In transfected 293T cells, GST–MEI-1 associated with MEL-26, but not with the related MATH-BTB protein F37A4.9 expressed at similar levels (Fig. 3b). Moreover, bacterial GST–MEI-1 was found to interact with *in vitro*-translated MEL-26 (Fig. 3c). By contrast, MEI-2 did not associate with MEL-26 *in vitro* (Fig. 3c) or in the two-hybrid system (data not shown). Thus, it is likely that MEL-26 and MEI-1 interact directly.

Genetically identified alleles of interacting proteins can frequently provide insight into the biochemical anatomy of signalling processes. In *C. elegans*, mutation of C94 to tyrosine in the MATH domain or partial deletion of the BTB domain in MEL-26 leads to persistent MEI-1 expression in the first cell cycle¹² (Fig. 4a). Previous studies have defined the structure of the MATH domain in the tumour necrosis factor receptor component TRAF2, which interacts with a peptide epitope derived from CD40 through a groove generated by a twisted β -sheet²¹ (Fig. 4b). We modelled the MEL-26 sequence on the MATH domain from TRAF2 (Fig. 4c). Notably, the position of C94 in the MEL-26 model corresponds to F410, located in the peptide-binding groove, suggesting that it might be directly involved in recognition of CUL-3 targets such as MEI-1. Consistent with this, we found that MEL-26 C94Y failed to interact efficiently with MEI-1 in both the two-hybrid system and in transfected 293T cells (Fig. 4e, f), but maintained interaction with CUL-3 (Figs 2h and 4e, f).

Mutation of P99 to leucine in MEI-1 (Fig. 4a) also gives persistent MEI-1 expression in embryos¹³. This defect in MEI-1 seems to reflect an inability to associate with MEL-26, as MEI-1 P99L failed

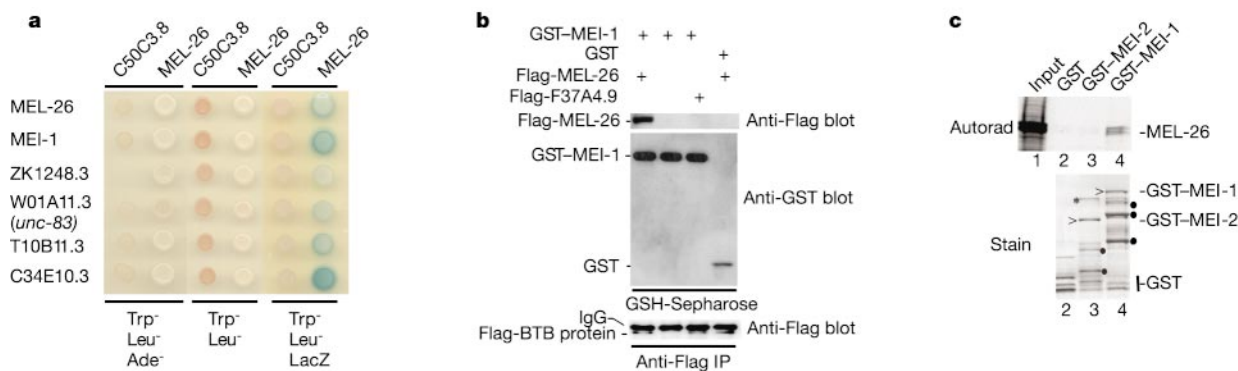


Figure 3 MEL-26 associates with its genetic target, MEI-1. **a**, Identification of MEL-26 interaction proteins using the two-hybrid system. Complementary DNA-library-derived pAD plasmids interact with pDB-MEL-26 but not with pDB-C50C3.8, a related MATH-BTB protein. **b**, GST–MEI-1 interacts with Flag-MEL-26 (lane 1) but not with Flag-C37A4.9 (lane 3) in transfected 293T cells. **c**, MEL-26 interacts with GST–MEI-1 *in vitro*. Two

micrograms of bacterial GST–MEI-1, GST–MEI-2 and GST (lower panel) proteins were used for binding with *in vitro*-translated MEL-26. The positions of full-length GST–MEI-1 and GST–MEI-2 proteins are shown, as is the position of GST fusion protein breakdown products (filled circles) and a nonspecific protein (asterisk) associated with GST–MEI-1 and GST–MEI-2.

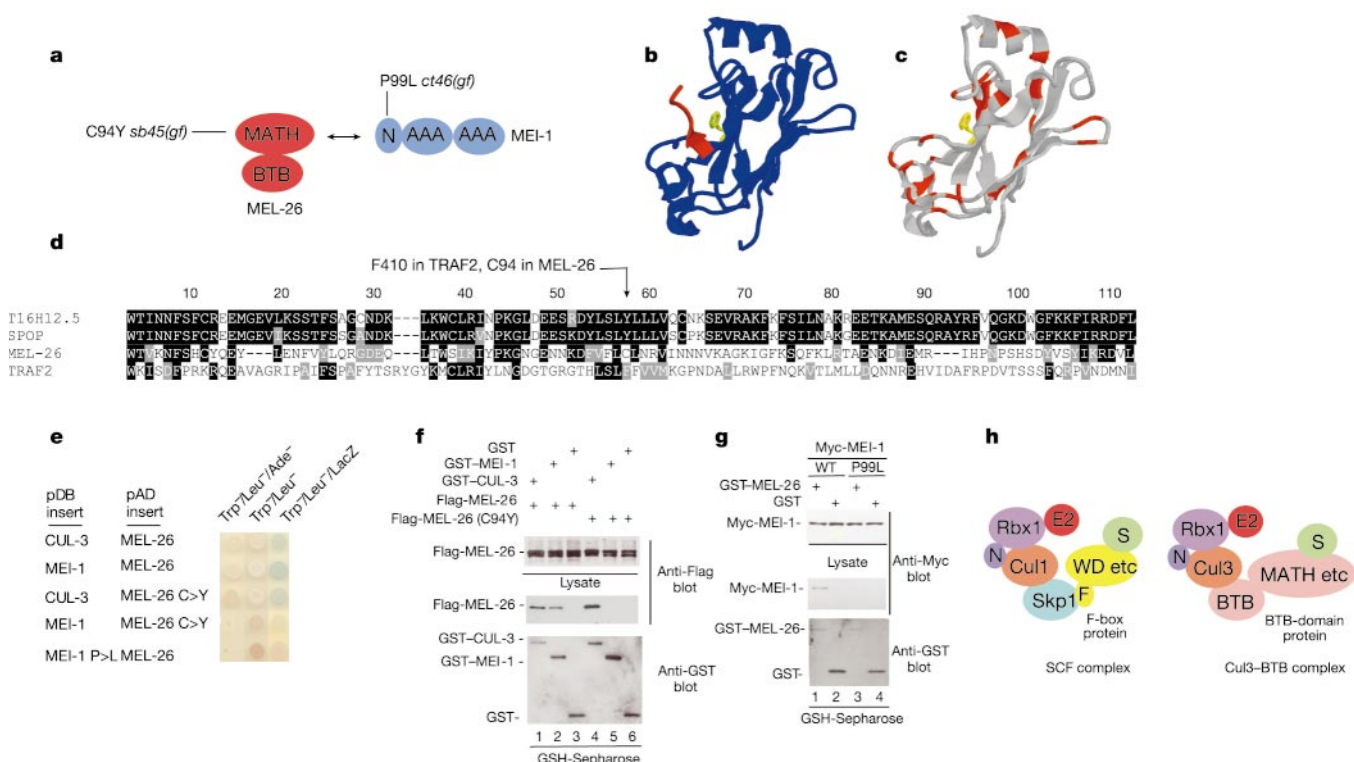


Figure 4 MEL-26 interacts with the N terminus of MEI-1 through its MATH domain. **a**, Domains in MEL-26 and MEI-1 and the locations of genetically defined *mel-26* and *mei-1* alleles^{12,13}, which give rise to defects in MEI-1 elimination at the meiosis–mitosis transition. **b**, Structure²¹ of the MATH domain from TRAF2 (blue) bound to a peptide from CD40 (red). **c**, Residues conserved between TRAF2 and MEL-26 (red) mapped onto the TRAF2 MATH domain. The position of Phe 410 (analogous to Cys 94 in MEL-26) is shown in yellow. **d**, Clustal W alignment of the MATH domains of human SPOP, *C. elegans*

T16H12.5 and MEL-26, and human TRAF2. **e–g**, The MEI-1–MEL-26 interaction is abolished by genetically identified mutations in the MATH domain of MEL-26 (C94Y, *sb45(gf)*) or the N-terminal domain of MEI-1 (P99L, *ct46(gf)*) in both the two-hybrid system (**e**) and in transfected 293T cells (**f, g**). **h**, Schematic representation of SCF and Cul3–BTB (SCF3) complexes emphasizing the notion that BTB-domain proteins merge the functional properties of Skp1 and F-box proteins into a single polypeptide specificity factor for CUL-3. S, substrate.

to interact with MEL-26 in the two-hybrid system (Fig. 4e) and the MEI-1–MEL-26 interaction was substantially diminished in transfected 293T cells (Fig. 4g). The coherence between the biochemical interactions examined here and the previously described genetic interactions indicate that MEL-26 functions directly in the elimination of MEI-1 through recruitment of the CUL-3 ubiquitin ligase. MEL-26 interacts with CUL-3 through its BTB domain and with the N terminus of MEI-1 through its MATH domain. A direct test of this hypothesis requires reconstitution of MEI-1 ubiquitination by a MEL-26–CUL-3 complex, which so far has been unsuccessful, possibly owing to the absence of other essential cofactors.

The use of modular adaptor proteins such as F-box and BC-box proteins as substrate-specific receptors provides a means to target diverse substrates to a common core ubiquitin ligase composed of Rbx1 and either Cul1 or Cul2/Cul5, respectively. This study together with the accompanying paper¹⁹ identifies a family of BTB-domain proteins as substrate-specific receptor modules for CUL-3-based E3 ligases. Evidence for this comes from our biochemical analysis of MEL-26 and its target MEI-1, wherein the biochemical properties we have defined for formation of MEI-1–MEL-26–CUL-3 complexes mirrors the genetic requirements for elimination of MEI-1 during the first mitotic division. However, unlike F-box and BC-box proteins, which interact with cullins through the adaptor proteins Skp1 and elongin C, BTB proteins appear to associate directly with CUL-3. Thus, BTB proteins merge the functional properties of Skp1 and F-box proteins into a single polypeptide (Fig. 4h).

BTB proteins contain a variety of putative protein interaction

domains, including MATH domains (38 in *C. elegans*, 2 in humans), WD40 repeats (1 in *C. elegans*, 2 in humans), Zn finger repeats (2 in *C. elegans*, 43 in humans) and kelch repeats (6 in *C. elegans*, 56 in humans). We have found that human BTB proteins containing MATH and kelch domains, as well as BTB-only proteins, interact with human CUL3 (our own unpublished data), indicating that this system extends to mammals. Although most BTB-domain proteins contain additional known protein interaction domains, approximately 30% lack recognizable domains and may have specialized protein interaction domains to recruit substrates or be substrates themselves, like some F-box proteins²². Given the number of F-box, BC-box and BTB proteins in eukaryotes (>300 in humans and >800 in *C. elegans*), SCF-like complexes seem to have the potential to control an extremely large number of signalling events through the ubiquitin–proteasome pathway. Although initially used to designate Skp1–Cul1–F-box protein complexes, the term SCF can be used conceptually in the description of a larger collection of cullin-based E3 ligases. Each SCF-like complex contains, in addition to the common catalytic core, a modular substrate-specific receptor that possesses structural homology with Skp1 and functional homology with C-terminal substrate interaction domains of F-box proteins. In addition, some SCF complexes contain further components—for example, elongin B in the case of Cul2/5 complexes. For purposes of nomenclature simplicity and to emphasize the conceptual similarity among cullin-based E3, we would like to propose that the different SCF-related complexes be referred to as SCF1, SCF2, SCF3, and so on, where the number refers to the identity of the cullin component. □

Methods

Plasmid construction and library screening

Plasmids expressing the indicated genes were generated by recombination of pENTR plasmids (Invitrogen) with one of the following pDEST plasmids: pDEST-AD (Gal4 activation domain), pDEST-DB (Gal4 DNA-binding domain), pCMV-Myc-DEST, pCMV-Flag-DEST, pCMV-GST-DEST, pT7-His6-DEST and pT7-GST-DEST. Two hybrid screens were performed using *Saccharomyces cerevisiae* strain PJ49-4A and either a *C. elegans* cDNA library or an 'ORFEOME' library containing about 12,000 *C. elegans* open reading frames^{14,23}. Selection was accomplished on synthetic complete medium lacking tryptophan, leucine and adenine for 3–5 days at 30 °C. Colonies were tested for LacZ activity using an overlay assay. Mutations were generated using a Gene Editor system (Promega).

Protein interactions

Protein interactions were examined in 293T cells 36 h after Fugene-6-mediated (Roche) transfection of the indicated plasmids. Cells were lysed in 0.1% Nonidet P40 buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA and protease inhibitors (Roche)), and cleared by centrifugation (14,000g, 15 min) before immunoprecipitation. Immune complexes were subjected to SDS-PAGE and immunoblotting using the indicated antibodies. For *in vitro* binding, immobilized GST fusion proteins (made in BL21/DE3 cells) were incubated with ³⁵S-methionine-labelled MEL-26. Washed beads were subjected to SDS-PAGE and MEL-26 was visualized by autoradiography. To examine MEL-26–CUL-3 interactions in *C. elegans*, 0.5 mg embryonic extract was used for immunoprecipitation using anti-MEL-26 antibodies. Immune complexes were immunoblotted with anti-CUL-3 and anti-NED-8 antibodies. Anti-MEL-26 and anti-CUL-3 were provided by M. Peter.

Received 15 July; accepted 14 August 2003; doi:10.1038/nature01985.
Published online 3 September 2003.

1. Koepp, D. M., Harper, J. W. & Elledge, S. J. How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell* **97**, 431–434 (1999).
2. Deshaies, R. J. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* **15**, 435–467 (1999).
3. Kipreos, E. T., Lander, L. E., Wing, J. P., He, W. W. & Hedgecock, E. M. *cul-1* is required for cell cycle exit in *C. elegans* and identifies a novel gene family. *Cell* **85**, 829–839 (1996).
4. Zheng, N. *et al.* Structure of the Cul1-Rbx1-Skp1-F box^{Skp2} SCF ubiquitin ligase complex. *Nature* **416**, 703–709 (2002).
5. Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219 (1997).
6. Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**, 209–219 (1997).
7. Bai, C. *et al.* Skp1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263–274 (1999).

8. Singer, J. D., Gurian-West, M., Clurman, B. & Roberts, J. M. Cullin-3 targets cyclin E for ubiquitination and controls S phase in mammalian cells. *Genes Dev.* **13**, 2375–2387 (1999).
9. Winston, J. T., Chu, C. & Harper, J. W. Culprits in the degradation of cyclin E apprehended. *Genes Dev.* **13**, 2751–2757 (1999).
10. Kurz, T. *et al.* Cytoskeletal regulation by the Nedd8 ubiquitin-like protein modification pathway. *Science* **295**, 1294–1298 (2002).
11. Pintard, L. *et al.* Neddylation and deneddylation of CUL-3 is required to target MEL-1/Katanin for degradation at the meiosis-to-mitosis transition in *C. elegans*. *Curr. Biol.* **13**, 911–921 (2003).
12. Dow, M. R. & Mains, P. E. Genetic and molecular characterization of the *Caenorhabditis elegans* gene, *mel-26*, a postmeiotic negative regulator of mei-1, a meiotic-specific spindle component. *Genetics* **150**, 119–128 (1998).
13. Clark-Maguire, S. & Mains, P. E. *mei-1*, a gene required for meiotic spindle formation in *Caenorhabditis elegans*, is a member of a family of ATPases. *Genetics* **136**, 533–546 (1994).
14. Reboul, J. *et al.* *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nature Genet.* **34**, 35–41 (2003).
15. Collins, T., Stone, J. R. & Williams, A. J. All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol. Cell. Biol.* **21**, 3609–3615 (2001).
16. Schulman, B. A. *et al.* Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex. *Nature* **408**, 381–386 (2000).
17. Aravind, L. & Koonin, E. V. Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain. *J. Mol. Biol.* **285**, 1353–1361 (1999).
18. Uren, A. G. & Vaux, D. L. TRAF proteins and meprins share a conserved domain. *Trends Biochem.* **21**, 244–245 (1996).
19. Pintard, L. *et al.* The BTB protein MEL-26 functions as a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. *Nature* advance online publication; 3 September 2003 (doi:10.1038/nature01959).
20. Srayko, M., Buster, D. W., Bazirgan, O. A., McNally, F. J. & Mains, P. E. MEL-1/MEL-2 katanin-like microtubule severing activity is required for *Caenorhabditis elegans* meiosis. *Genes Dev.* **14**, 1072–1084 (2000).
21. Ye, H., Park, Y. C., Kreishman, M., Kieff, E. & Wu, H. The structural basis for the recognition of diverse receptor sequences by TRAF2. *Mol. Cell* **4**, 321–330 (1991).
22. Zhou, P. & Howley, P. Ubiquitination and degradation of the substrate recognition subunits of SCF ubiquitin-protein ligases. *Mol. Cell* **2**, 571–580 (1998).
23. Walhout, A. J. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122 (2000).

Acknowledgements We thank M. Peter for anti-CUL-3 and anti-MEL-26 antibodies. This work was supported by an NIH grant (to J.W.H. and S.J.E.), the Welch Foundation (to J.W.H.), and the National Human Genome Research Institute (to M.V.). Y.W. was supported by a Department of Defense predoctoral fellowship. S.J.E. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.W.H. (wade_harper@hms.harvard.edu).

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Netherlands/ Italy/ Spain/ Portugal/ Belgium:

Evelina Rubio Hakansson (4973)
Scandinavia: Sille Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Natureevents:

Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakakamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

Working for peanuts

In the United States, student loans have been a fact of life for years, but in Britain they are a relatively recent phenomenon. In both countries, the process can have subtle and not-so-subtle effects on a student's career path — especially when you consider the final debt load versus the salary potential of the course being studied.

Some US analysts say that many students who go on to postgraduate courses opt for law or business administration rather than science because of the greater salary potential. And in medicine, similar concerns have led to a dearth of physician scientists — those who focus on research rather than clinical care. Postgraduates with such dual skills must choose between practice, where higher salaries can dig them out of debt, and research, where they will be mired in loan repayments for longer.

In the United States, there have been some attempts to redress the situation — for example, there are some loan-repayment programmes for MD/PhDs who choose research over practice. Meanwhile, in Britain, one graduate has protested about the size of his loan by pushing a peanut with his nose seven miles through the streets of London, from Goldsmith College to 10 Downing Street. Although different in dramatic appeal, both attempts at loan forgiveness have limited efficacy. The US schemes tend to target specialized areas of research, usually leaving basic-science loans unforgiven. As for the peanut pusher, Prime Minister Tony Blair has already said that he will not write off Mark McGowan's £15,000 (US\$24,000) debt.

Perhaps what is needed is an increase in broad loan-forgiveness programmes for people going into research areas where there is a deficit of skilled workers. Or maybe debt-beleaguered graduate students on both sides of the Atlantic could orchestrate a mass peanut push in the hope of achieving such a change.

Paul Smaglik
Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Growing opportunities
for US biomedical
engineering p324

Europe takes on the
training load p326

Room for entrepreneurs p327

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Biomedicine meets engineering

Steady philanthropic support, new federal funding and increased commercial interest are mixing in the United States to create a large and growing number of opportunities for a new breed of multidisciplinary researcher, says Virginia Gewin.

In today's depressed job market, Stephanie Wandke can't believe that she is turning away multiple job offers. Like many students attracted to biomedical engineering, Wandke wanted to put her maths and science aptitude to use by making a tangible difference to patients' lives. After leaving Marquette University in Milwaukee, Wisconsin, two years ago, she quickly found work in industry developing less invasive forms of surgery using a robotic arm. When corporate calamity hit her first employer, Wandke quickly found herself swamped with job offers, from which she picked a post as a clinical specialist/sales associate at I-Flow, a company in Lake Forest, California, that specializes in state-of-the-art anaesthesia delivery.

Wandke is not alone. Job prospects are bright in biomedical engineering (BME), which combines engineering design skills with biological expertise, and students are flocking to enter the field. According to the US Department of Labor, the number of BME jobs in the United States is expected to increase by 31.4% during the next seven years, more than double the average predicted rate in other fields.

Opportunities are arising on several fronts. The recently established National Institute of Biomedical Imaging and Bioengineering (NIBIB) is generating training and research schemes. New university departments are forming and are recruiting researchers to train the onslaught of applicants. And the biomedical industry's interest in engineering is growing.

ENGINEERING A NEW INSTITUTE

Set up in 2001, the NIBIB is the newest member of the US National Institutes of Health (NIH) and already boasts a \$280-million budget. In 2002, more than \$100 million went on nearly 300 extramural grants. It offers research opportunities in everything from small-animal imaging to tissue engineering and neuroinformatics.

Beyond the NIBIB, BME continues to flourish throughout the NIH. The bioengineering consortium BECON — a collection of representatives from all of the NIH institutes, centres and other interested federal agencies — coordinates \$900-million worth of research and training opportunities throughout the NIH. And the National Science Foundation (NSF) is

also getting into the BME act with its integrative graduate education and research traineeship (IGERT) programme.

Taking a cue from Europe (see page 326), the biggest growth in training programmes is occurring at the undergraduate level. The NIH and NSF are promoting interest in the field through summer programmes. Undergraduates and first-year graduate students can explore the field through joint NIH/NSF-funded bioengineering and bioinformatics summer institutes at nine universities around the country, or come to the NIH itself as part of the BME summer internship programme.

CHANGING LANDSCAPE

Ironically, perhaps the single biggest player in establishing the NIBIB is closing its doors in 2007. More than 20 years ago, the Whitaker Foundation began providing what would amount to \$720 million to establish biomedical engineering departments at top universities. "The Whitaker Foundation has had more of an impact on a single profession than any other foundation I've ever heard of," says Pat Horner, current president of the Biomedical Engineering Society (BMES).

The figures speak for themselves. The total number of undergraduate BME programmes has doubled in the past five years. Experts anticipate that the number of accredited departments will also increase — from 24 to 50 in the short term. During the past ten years, BME enrolment has doubled while engineering undergraduate enrolment has essentially been flat, says Bruce Hamilton, NSF division director for bioengineering and environmental systems.

BME departments enjoy an unusual luxury: a plethora of coveted university positions. At Boston University, one of the oldest BME departments in the United States will increase its faculty from 23 members to more than 30 following a substantial gift from Whitaker, making it the largest in the nation. Competition for top students and faculty has reached fever pitch.

The competition for research dollars has escalated as well. "We are basically being swamped with proposals from junior-level faculty members right now," says Hamilton. The NSF's contribution has



Biomedical engineering in action: these students are examining a heart catheterization procedure.

grown to some \$30 million. "The funding is growing," says Hamilton, "but the proposal pressure is growing even faster."

Industry too is taking its share of talent from the growing pool of expertise (see page 327). "The most exciting thing is that industry is seeing that the quantitative computer-oriented approach to solving biomedical problems is essential," says Murray Sachs, chair of the BME department at Johns Hopkins University in Baltimore, Maryland. Indeed many industry leaders are harnessing this approach in their efforts to meet the predicted needs of ageing baby boomers.

"We're going to have to figure out how to take care of these people as they reach retirement age," says Dane Miller, president and chief executive of Biomet, a company based in Warsaw, Indiana, that develops total reconstructive joints. He adds that biomedical engineers are needed to help fine-tune a less invasive system of replacing joints.

Medical-device developers at Medtronic in Minneapolis, Minnesota, echo that trend. "We're always looking for someone with a biomedical-engineering slant," says Len Twetan, Medtronic's senior product development manager.

MARKETABLE QUALITIES

A key quality of a successful biomedical engineer is the ability to understand and communicate both engineering and medical jargon. "We tend to see a much different recruit out of biomedical engineers than we saw a decade ago," says Jerry Fisher, vice-president of research and technical services at medical manufacturer Baxter, based in Deerfield, Illinois.



Growth industry: demand for biomedical engineering is on the rise according to Bruce Hamilton.

"Over the next five years or so, we will replace the more traditional engineering roles with biomedical engineers."

Now that medical-device manufacturers have found biomedical engineers, the BMES is mounting a campaign to take the pharmaceutical industry by storm. "The drug industry is beginning to see the value of biomedical engineers," says Sachs. "Biomedical engineers are unique because they can move bioinformatics and computational biology forwards."

Mining the human genome and proteome for life-saving advances looks to be a good way to ensure long-term job security. Wandke sums up the BME career prospects nicely: "Biomedical engineering is made up of two fields that will never go away, are constantly changing, and are in demand," she says. "It's an awesome place to position yourself."

Although competition for top faculty positions can sometimes be fierce, the trend in biomedical engineering has now trickled down to graduate student stipends. In response to the number of top-quality students being drawn to the field, some BME departments are offering as much as \$10,000 over the average \$18,000 NIH predoctoral stipend. "We've heard of some extraordinarily high stipends offered to graduate students," says Kenneth Lutchien, chair of Boston University's BME department. "We view this as more destructive than constructive because it is difficult to predict who the superstars will be four to five years later," he adds.

Sachs agrees. "Ten years ago, we were one of the only games in town and we got most of the graduate students that we wanted," he says. "Now it has become considerably harder."

Virginia Gewin is a freelance science writer in Corvallis, Oregon.

Web links

National Institute of Biomedical Imaging and Bioengineering
 ♦ www.nibib.nih.gov
 NIH/NSF Training Opportunities
 ♦ www.nibib1.nih.gov/training/NIHbioe.html
 NIH Bioengineering Consortium (BECON)
 ♦ www.becon1.nih.gov/becon_funding.htm
 Biomedical-engineering summer internship programme at the NIH
 ♦ www.nih.gov/od/ors/dbeps/besip/index.htm
 NSF integrative graduate education and research traineeship
 ♦ www.nsf.gov/home/crssprgm/igert/start.htm
 NIH/NSF-funded bioengineering and bioinformatics summer institutes
 ♦ bbi.eecom.com
 Biomedical Engineering Society
 ♦ www.bmes.org

Europe chips in for training

The United States may have more coordinated funding, but Europe is taking the lead in training biomedical engineers. Ralf Jox reports.

Almost 20 years ago, Peter Fromherz, a biophysicist at the Max Planck Institute of Biochemistry in Munich, Germany, connected living cells to microchips to measure the electrical signals at the interface between the two. With sensor-chip capacity steadily increasing, he has since made significant progress in getting neurons and silicon chips to communicate with each other.

Meanwhile, the chip industry has started to catch up with what once sounded like science fiction, by exploring the commercial opportunities the nascent field offers. The German semiconductor company Infineon Technologies based in Munich, for example, has invested several million euros in the production of a prototype super-chip, tailor-made for Fromherz's research.

Industry's investment is fuelling a growth in biomedical engineering (BME), an emerging discipline at the crossroads of engineering, life sciences and health care. BME research institutes and training courses have sprung up rapidly during the past ten years. Although the United States may be leading the field with its National Institute of Biomedical Imaging and Bioengineering (NIBIB; see page 324), BME is growing in Europe too. Universities throughout the European Union (EU) have responded by developing innovative training programmes at all levels.

Although Europe has no special BME funding body comparable to NIBIB, Ferdinando Grandori, the director of the Italian Institute of Biomedical Engineering in Milan, says that European funding opportunities for biomedical-engineering projects are "excellent".

COLLECTIVE FORCE

Joachim Nagel, head of the BME department at the University of Stuttgart in Germany, would like to see a more centralized approach to BME research and training in Europe. The creation of the European Alliance for Medical and Biological Engineering and Science (EAMBES) is a start. This umbrella organization brings together numerous national scientific societies and academic institutions, with the goal of promoting medical and biological engineering at European and national levels.

Nagel has compiled for EAMBES profiles of BME courses offered at more than 200 universities in 28 European countries. His goal is to create a European-



Bridging the gap: a neurochip, based on the work of Peter Fromherz, links biology and technology together.

wide accreditation system so courses in different countries can be both comparable and competitive.

Academic training of biomedical engineers in Europe has long been regarded as a postgraduate specialism. But this is changing with the recognition that the next wave of pioneers must get an early start, says Jos Vander Sloten, a bioengineer at the Catholic University of Leuven in Belgium, and secretary-general of EAMBES. "For young scientists to succeed in this field, the best way is to combine biology and engineering science right from the beginning of their studies," he says.

In 1997, Eindhoven and Maastricht universities in the Netherlands established undergraduate BME courses. Other universities throughout Europe are now following their lead. Some of the first students to take those programmes are now into graduate work — and leaving some of their mentors behind. "They are so well trained that some professors find it hard to grasp what their research is about," says Fons Sauren, a biomedical engineer who oversees the University of Eindhoven's BME programme. PhD programmes and postgraduate masters courses are also proliferating.

Fromherz views the budding euphoria with caution. "My work on neurochips has stirred incredible public interest," he says. "But I'd rather carry on working than bask in the hype." And there he goes, listening again to the slowly rising whispers between neurons and silicon chips. ■

Ralf Jox is a biomedical scientist and former intern in *Nature's* Munich office.

INFINEON

Engineering your own path

Prospects are good for biomedical engineers across industry — and there's still room for entrepreneurs, say Ralf Jox and Virginia Gewin.

Although the creation of new institutes and organizations has justifiably gained a lot of press, they are not yet providing many biomedical engineering (BME) jobs. Bodies such as the US National Institute of Biomedical Imaging and Bioengineering (NIBIB) and the European Alliance for Medical and Biological Engineering and Science are expected to create many new jobs in the future. Meanwhile, most graduates find their place in industry.

Europe has about 350,000 biomedical engineers, almost as many as the United States, according to EUCOMED, an association representing European medical-technology companies. The European market for BME, with an annual turnover of €45 billion (US\$50 billion), is led by Germany, Britain and France.

"In industry, biomedical engineers work in corporate research, production, quality control and management," says Katharina Jäger from the human-resources department of Baxter, a global medical-technology company based in Vienna. "University graduates usually acquire leading positions after one or two years, and salaries rise from around €35,000 to €50,000," she says.

Job opportunities in industry will remain good, says Arjen Schat, European recruitment manager of Philips Medical Systems in Eindhoven, the Netherlands. "In the coming years we will need a gradually increasing number of biomedical engineers."

The only European survey so far about career paths of biomedical engineers, conducted by Jaakko Malmivuori at Tampere University of Technology in Finland, yielded a rosy picture. "Eighty per cent of our graduates from 1979 to 1997 were employed in full-time jobs, and one-third moved quickly to managerial positions," he says.

SMALL BUSINESS, BIG OPPORTUNITY

Although jobs in the US biomedical industry have historically been in larger firms such as Medtronic or Baxter, the growth in industrial posts is likely to come predominantly from smaller spin-offs. Indeed, entrepreneurial fever has hit the US biomedical-engineering community. Kyriacos Athanasiou, president-elect of the Biomedical Engineering Society, has himself started four businesses during the past decade.

Athanasiou set up two of his businesses with nothing but venture-capital funding. Over the past few years, as venture capital in the United States dried up, Athanasiou turned to government grants. In the United States, that meant Small Business Innovation



Opportunity knocks: bioengineer Kyriacos Athanasiou has started up four companies of his own in recent years.

Research (SBIR) grants. These provide money through government agencies such as the National Institutes of Health (NIH) to help develop new ideas.

Although there is no measure of how many SBIR grants were devoted to BME applications before the creation of the NIBIB, the innovative spirit that the scheme fosters is evident. Only 13% of applications to the NIBIB in 2002 were for SBIR grants, says Joan Harmon, director of the institute's office of extramural policy. But in 2003, that rose to more than 32%. Currently, funds available for SBIR support comprise 3.4% of the NIBIB's \$280-million budget.

Robert Schmidt, biomedical engineer and entrepreneur, admits that the ultimate pay-off from SBIR companies can be a long time in coming. Following his SBIR funding, Schmidt invested \$10,000 in his fledgling Cleveland Medical Devices — a company that develops physiological monitoring equipment. After six years with no salary, he now has \$3 million in sales a year.

As NIH funding continues to grow, the trend is for grants to be larger rather than more numerous. More than \$1.6 billion of the NIH budget now goes into SBIR. As the NIBIB becomes more firmly established, it is anticipated that funding will continue to increase. "The money follows the talent, and that's increasingly with small companies," says Ann Eskesen, president of the Innovation Development Institute, a leading SBIR think-tank and programme advocate.

Ralf Jox is a biomedical scientist and former intern in Nature's Munich office; Virginia Gewin is a freelance science writer in Corvallis, Oregon.

Web links

Baxter
 ♦ www.baxter.com
 Philips Medical Systems
 ♦ www.medical.philips.com
 Tampere University of Technology,
 Ragnar Granit Institute
 ♦ ee.tut.fi/rgi
 Tampere University of Technology,
 biomedical-engineering survey
 ♦ www.ee.tut.fi/rgi/bme-survey
 EUCOMED Medical Technology
 ♦ www.eucomed.be
 NIH Small Business Funding
 Opportunities
 ♦ grants1.nih.gov/grants/funding/sbir.htm